BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Comparison of epsilon- and delta-class glutathione *S*-transferases: the crystal structures of the glutathione *S*-transferases DmGSTE6 and DmGSTE7 from *Drosophila melanogaster*

Michele Scian,[a] Isolde Le Trong,[b,c] Aslam M. A. Mazari,[d] Bengt Mannervik,[d] William M. Atkins[a] and Ronald E. Stenkamp[b,c,e]*

[a]Department of Medicinal Chemistry, University of Washington, Box 357610, Seattle, WA 98195-7610, USA, [b]Department of Biological Structure, University of Washington, Box 357420, Seattle, WA 98195-7420, USA, [c]Biomolecular Structure Center, University of Washington, Box 357742, Seattle, WA 98195-7742, USA, [d]Department of Neurochemistry, Arrhenius Laboratories, Stockholm University, SE-10 691 Stockholm, Sweden, and [e]Department of Biochemistry, University of Washington, Box 357430, Seattle, WA 98195-7430, USA. *Correspondence e-mail: stenkamp@u.washington.edu

Received 25 June 2015
Accepted 22 July 2015

Edited by R. McKenna, University of Florida, USA

Keywords: glutathione *S*-transferase; detoxication; glutathione binding.

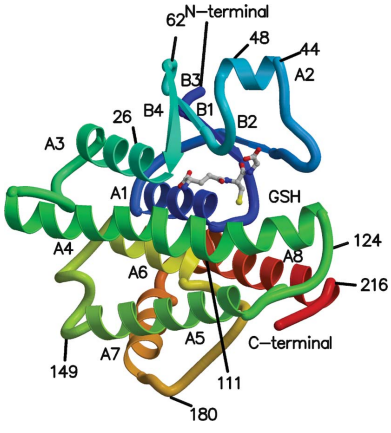
PDB references: DmGSTE6, 4pnf; DmGSTE7, 4png

Supporting information: this article has supporting information at journals.iucr.org/d

Cytosolic glutathione transferases (GSTs) comprise a large family of enzymes with canonical structures that diverge functionally and structurally among mammals, invertebrates and plants. Whereas mammalian GSTs have been characterized extensively with regard to their structure and function, invertebrate GSTs remain relatively unstudied. The invertebrate GSTs do, however, represent potentially important drug targets for infectious diseases and agricultural applications. In addition, it is essential to fully understand the structure and function of invertebrate GSTs, which play important roles in basic biological processes. Invertebrates harbor delta- and epsilon-class GSTs, which are not found in other organisms. *Drosophila melanogaster* GSTs (DmGSTs) are likely to contribute to detoxication or antioxidative stress during development, but they have not been fully characterized. Here, the structures of two epsilon-class GSTs from *Drosophila*, DmGSTE6 and DmGSTE7, are reported at 2.1 and 1.5 Å resolution, respectively, and are compared with other GSTs to identify structural features that might correlate with their biological functions. The structures of DmGSTE6 and DmGSTE7 are remarkably similar; the structures do not reveal obvious sources of the minor functional differences that have been observed. The main structural difference between the epsilon- and delta-class GSTs is the longer helix (A8) at the C-termini of the epsilon-class enzymes.

1. Introduction

The cytosolic glutathione transferases (GSTs) comprise a large family of detoxication enzymes with additional possible functions as chaperones and regulators of signal transduction (Armstrong, 1997; Hayes *et al.*, 2005; Zhang *et al.*, 2014). GSTs provide a critical line of defense against chemical insult, oxidative stress and possibly cancer (Sharma *et al.*, 2004; Henderson & Wolf, 2005). Several classes of cytosolic GSTs, designated by Greek letter names with corresponding alpha-numeric letter designations for the proteins that they encode, are represented in a wide range of species, including the alpha (A), mu (M), pi (P), delta (D), epsilon (E), zeta (Z), theta (T) and omega (O) classes. Although their main catalytic function results in conjugation of the tripeptide glutathione (GSH) to electrophilic xenobiotics, GSTs also catalyze double-bond isomerization without net consumption of GSH (Tars *et al.*, 2010). The cytosolic GSTs from mammals have been well characterized structurally and mechanistically, and many



aspects of their differential substrate specificities are well understood (Johnson *et al.*, 1993; Balogh *et al.*, 2009; Norrgård *et al.*, 2006; Tars *et al.*, 2006, 2010). For example, the molecular basis for the substrate selectivities of GSTs A4-4, M1-1, P1-1 and T1-1 is clearly related to structural differences in the 'H-site,' which is the binding site for hydrophobic xenobiotics or the products of oxidative stress, which lies adjacent to the highly conserved glutathione-binding site (G-site).

In addition to the GST isoforms expressed in human tissues, epsilon-class GSTs from insects and other GST classes from parasites are important in human health and agriculture because they are potential pharmacological targets. For example, GSTs from *Plasmodium falciparum* are likely to be responsible for the clearance of, and resistance to, several antimalarial drugs (Deponte & Becker, 2005; Mahajan & Atkins, 2005; Mohring *et al.*, 2014). Alternatively, the parasite GSTs provide a critical role in heme clearance and avoidance of oxidative stress, so their selective inhibition could result in an effective malarial therapy (Harwaldt *et al.*, 2002). Similarly, the malarial mosquito vector *Anopheles sinensis* has been difficult to eradicate, possibly owing to GST-mediated metabolism of insecticides, including DDT (Wang *et al.*, 2008; Reddy *et al.*, 2011; Jones *et al.*, 2012). Owing to their unique role in parasite or vector physiology, and their potential role in insecticide metabolism, GSTs from insects have received significant consideration as antimalarial and antitrypanosomal targets. Intriguingly, epsilon-class GSTs are not found in mammals, further amplifying their attractiveness as drug targets in infectious diseases. In contrast to human GSTs, the structural basis for substrate selectivity among epsilon-class GSTs is not well defined.

Among epsilon-class GSTs, isoforms from *Drosophila* may be uniquely interesting (Saisawang *et al.*, 2012). Studies from several models of aging, including *Drosophila*, indicate that oxidative stress is a critical determinant of life span. Gene-expression changes associated with aging or oxidative stress in *Drosophila* include the upregulation of GSTs (Dudas & Arking, 1995; McElwee *et al.*, 2007; Landis *et al.*, 2012), making them an interesting candidate marker or causal component of aging in this model (Landis *et al.*, 2012). Furthermore, all 41 GSTs of the *Drosophila* proteome are expressed in the late embryonic stage, suggesting their importance in development (Saisawang *et al.*, 2012). In order to further understand the structural basis for the substrate selectivity of GSTs from *D. melanogaster*, we have compared the structures of DmGSTE6 and DmGSTE7 with other epsilon- and delta-class GSTs. The results extend our understanding of class-dependent differences among canonical GST structures.

Table 1
Diffraction and refinement statistics.

Values in parentheses and square brackets are for the inner and outer shells, respectively.

	DmGSTE6	DmGSTE7
Diffraction statistics		
Unit-cell parameters (Å, °)	a = 52.57, b = 208.49, c = 86.45, $\alpha = \gamma = 90$, $\beta = 92.45$	a = 56.73, b = 87.09, c = 87.23, $\alpha = \beta = \gamma = 90$
Space group	$P2_1$	$P2_12_12_1$
Beamline	Beamline 11-1, SSRL	Beamline 11-1, SSRL
Low-resolution limit (Å)	39.90 (39.90) [2.22]	34.58 (34.58) [1.56]
High-resolution limit (Å)	2.11 (6.67) [2.11]	1.53 (8.38) [1.53]
R_{merge}	0.066 (0.018) [0.334]	0.046 (0.013) [0.372]
R_{meas}	0.079 (0.022) [0.412]	0.052 (0.015) [0.446]
$R_{p.i.m.}$	0.042 (0.012) [0.235]	0.024 (0.007) [0.238]
No. of observations	347743 (11674) [36497]	255608 (1948) [5129]
No. of unique reflections	102285 (3362) [12763]	63530 (444) [1779]
Mean $I/\sigma(I)$	16.2 (53.1) [3.3]	21.1 (82.1) [2.9]
Completeness (%)	96.2 (98.4) [82.7]	96.5 (95.4) [55.8]
Multiplicity	3.4 (3.5) [2.9]	4.5 (4.4) [2.9]
Refinement		
Resolution (Å)	39.9–2.1	61.6–1.5
No. of reflections, working set	92044	57011
No. of reflections, test set	5089	3223
R_{cryst}, all data	0.186	0.160
R_{cryst}, working set	0.183	0.159
R_{free}, test set	0.248	0.191
Bond r.m.s.d. (restrained) (Å)	0.014	0.010
Angle r.m.s.d. (restrained) (°)	1.65	1.43
No. of protein atoms	13985	3610
No. of heteroatoms	160	42
No. of solvent atoms	729	403
Ramachandran outliers as determined by *MolProbity* (%)	0.3	0.2
Ramachandran favored (%)	95.9	97.3
PDB code	4pnf	4png

2. Materials and methods

2.1. Protein expression and biophysical characterization

The genes for DmGSTE6 and DmGSTE7 were cloned into a pET vector as described elsewhere (Kjellander *et al.*, 2014), The proteins (with hexahistidine tags at their N-termini) were expressed in *Escherichia coli* strain BL21(DE3) and purified by affinity chromatography. Briefly, 6 × 500 ml *E .coli* cultures in Terrific Broth supplemented with 50 µg ml^{-1} kanamycin were incubated at 37°C and 200 rev min^{-1}. At an OD_{600} of 0.6, the cultures were induced with 1 m*M* IPTG and incubated for a further 6 h. The cells were then harvested by centrifugation (4000 rev min^{-1} for 1 h at 4°C), resuspended at ∼0.2 g ml^{-1} in 50 m*M* potassium phosphate (KP_i), 300 m*M* NaCl, 10 m*M* imidazole pH 7.4 and French-pressed three times at 13.8 MPa. After centrifugation for 30 min at 40 000 rev min^{-1} and 4°C, the pellets were discarded and the supernatants were loaded onto 10 ml (bed volume) His60 Ni Superflow resin (Clontech) equilibrated with 50 m*M* KP_i, 300 m*M* NaCl, 10 m*M* imidazole pH 7.4. After washing with two column volumes of 50 m*M* KP_i, 300 m*M* NaCl, 20 m*M* imidazole pH 7.4, the proteins were eluted with 50 m*M* KP_i, 300 m*M* imidazole pH 7.0. The protein-containing fractions were pooled, concentrated using

Table 2
Other GST structures discussed in this paper.

PDB code	GST class	Organism	Comments†	Identifier	Reference
4hi7	—	*Drosophila mojavensis*	GSH complex		Enzyme Function Initiative (unpublished work)
3vwx	Epsilon	*Musca domestica*	GSH complex	MdGST6B	Nakamura *et al.* (2013)
2il3	Epsilon	*Anopheles gambiae*		AgGSTE2	Wang *et al.* (2008)
2imi	Epsilon	*Anopheles gambiae*	GSH complex	AdGST1-3‡	Oakley *et al.* (2001)
2imk	Epsilon	*Anopheles gambiae*	GTX complex	AgGSTE2	Wang *et al.* (2008)
1jlv	Delta	*Anopheles dirus*	GSH complex	AdGST1-3‡	Oakley *et al.* (2001)
1v2a	Delta	*Anopheles dirus*	GTS complex	AdGST1-6	Udomsinprasert *et al.* (2005)
1r5a	Delta	*Anopheles dirus*	GTS complex	AdGSTD5-5	Udomsinprasert *et al.* (2005)
1pn9	Delta	*Anopheles gambiae*	GTX complex	AgGSTD1-6‡	Chen *et al.* (2003)
3ein	Delta	*Drosophila melanogaster*	GSH complex	DmGSTD1	Low *et al.* (2010)
3f6f	Delta	*Drosophila melanogaster*	Ligand-free?	DmGSTD10	Wongsantichon *et al.* (2012)
2fno		*Agrobacterium tumefaciens*	Ligand-free?		Kosloff *et al.* (2006)
1k3y	Alpha	*Homo sapiens*	GTX complex	HsGSTA1	Le Trong *et al.* (2002)

† GSH, glutathione; GTX, *S*-hexylglutathione; GTS, glutathione sulfonic acid. ‡ Alternate splice variants, not heterodimers.

an Amicon (Millipore) Ultra-15 centrifugal device (molecular-weight cutoff 10 000 Da) and dialyzed against 4 × 1 l 50 m*M* KP_i pH 7.0. A purity grade of >98% was then achieved by gel filtration using a Superdex 200 10/300 GL column (GE Healthcare; 10 × 300 mm, 13 µm particle size) with 10 m*M* Tris–HCl pH 7.0 running buffer (flow rate 0.50 ml min^{-1}; isocratic elution).

2.2. Crystallization and diffraction data collection

Crystals of DmGSTE6 in complex with GSH were obtained from hanging-drop vapor-diffusion experiments. The reservoir solution consisted of 0.1 *M* HEPES pH 7.5, 20%(*w*/*v*) PEG 4000, 10% 2-propanol, 2 m*M* GSH, 10 m*M* dithiothreitol (DTT). 2 µl drops of DmGSTE6 at 10 mg ml^{-1} in 10 m*M* Tris–HCl pH 7.0 were mixed with 2 µl drops of the reservoir solution and were allowed to equilibrate against the reservoir at room temperature. The space group of the resulting crystals is $P2_1$, with unit-cell parameters a = 52.57, b = 208.49, c = 86.49 Å, β = 92.45° and with four GST dimers (eight subunits) in the asymmetric unit.

The same experimental setup produced crystals of the GSH complex of DmGSTE7, but in this case the reservoir solution was 0.1 *M* Tris–HCl pH 7.5, 25%(*w*/*v*) PEG 4000, 2 m*M* GSH, 10 m*M* DTT. The protein solution contained DmGSTE7 at 10 mg ml^{-1} in 10 m*M* HEPES pH 7.0. Orthorhombic crystals grew over several months in space group $P2_12_12_1$, with unit-cell parameters a = 56.73, b = 87.09, c = 87.23 Å and with one GST dimer in the asymmetric unit.

Diffraction data for DmGSTE6 and DmGSTE7 were collected on beamline 11-1 at the Stanford Synchrotron Radiation Lightsource. The crystallization solutions provided sufficient cryoprotection for data collection at 100 K. The crystal-to-detector, exposure time, beam size, frame size and number of frames for DmGSTE6 were 450 mm, 2 s, 0.15 × 0.15 mm, 0.5° and 360, respectively, and those for DmGSTE7 were 300 mm, 0.2 s, 0.20 × 0.20 mm, 0.15° and 827, respectively. The data sets were processed using *XDS* (Kabsch, 2010). A summary of the statistics is presented in Table 1.

2.3. Structure solutions and refinements

The structure of DmGSTE6 was solved with *BALBES* (Long *et al.*, 2008) using Protein Data Bank (PDB) entries 2imi, 1jlv, 3ein, 3f6f and 1r5a (see Table 2) as template structures. Refinement of the model produced by *BALBES* made use of *REFMAC*5 (Murshudov *et al.*, 2011) in the *CCP*4 program suite (Winn *et al.*, 2011). 5% of the reflections were reserved for the calculation of R_{free} (Brünger, 1992).

XtalView (McRee, 1999) was used to view σ_A-weighted $|F_o| - |F_c|$ and $2|F_o| - |F_c|$ electron-density maps (Read, 1986; see Figs. 1 and 2) and to manipulate the molecular models. The structural models were evaluated during and after refinement using *MolProbity* (Chen *et al.*, 2010) and *ADIT* (Berman *et al.*, 2000).

A similar approach was used for DmGSTE7, but using *MOLREP* (Vagin & Teplyakov, 2010) in the *CCP*4 package and a DmGSTE6 dimer (chains *A* and *B*) as the test structure. The initial model had the DmGSTE6 amino-acid sequence, but difference electron-density maps were consistent with the

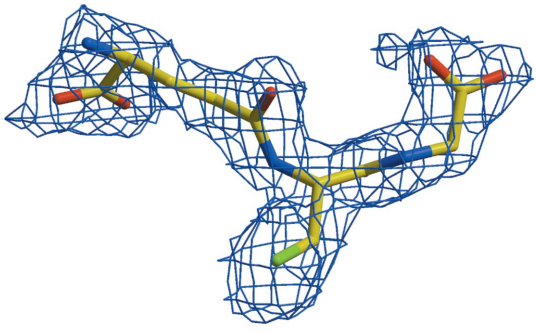


Figure 1
Final σ_A-weighted $2|F_o| - |F_c|$ electron density (Read, 1986) for the GSH molecule bound to subunit *A* of DmGSTE6 contoured at 1σ.

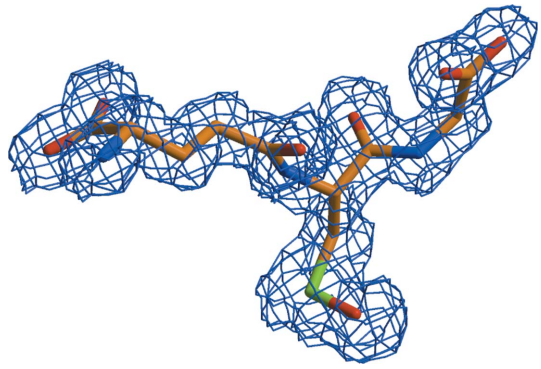


Figure 2
Final σ_A-weighted $2|F_o| - |F_c|$ electron density (Read, 1986) for the GSF molecule bound to subunit *A* of DmGSTE7 contoured at 1σ.

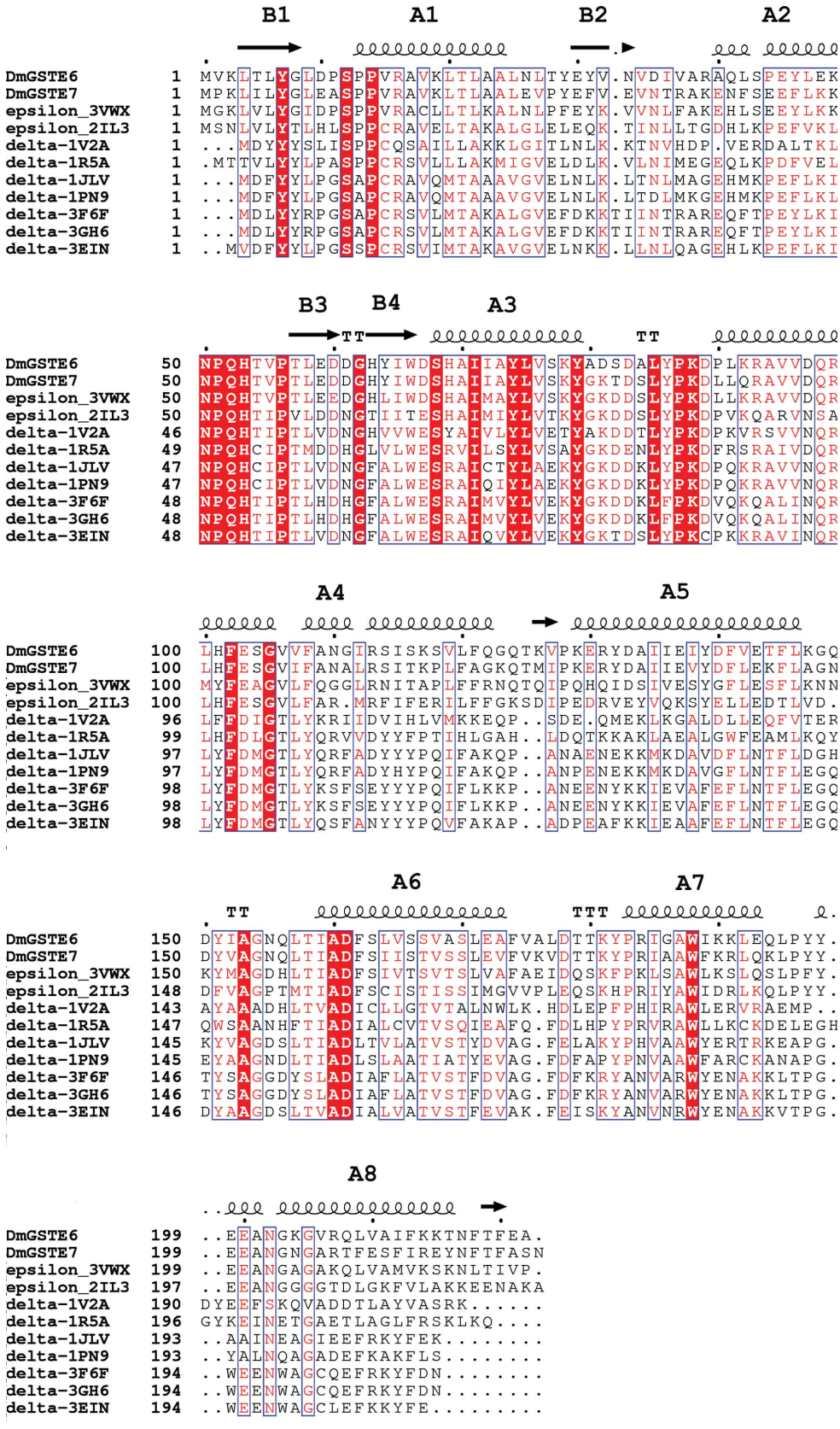


Figure 3
Sequence comparison of epsilon-class and delta-class GSTs. Residues highlighted in red are conserved. Residues shown in red type and surrounded by blue boxes are similar. The secondary-structure determination is for DmGSTE6.

sequence of DmGSTE7. (69% of the residues are identical in the two amino-acid sequences.) Appropriate amino-acid changes were made in the model, and refinement proceeded as for DmGSTE6. Although the crystals were grown in the presence of GSH, electron density for an O atom was found bound to the S atom of GSH in each of the protein subunits, so the structure of DmGSTE7 is that of its glutathione sulfenic acid complex (GSF). The oxidation products of cysteines are frequently observed in many crystal structures.

Amino-acid sequence alignments were obtained using *ClustalW* (http://www.ebi.ac.uk/Tools/msa/clustalw2/; Larkin *et al.*, 2007; Goujon *et al.*, 2010). Figs. 1 and 2 were generated using *XtalView* (McRee, 1999) and *Raster3D* (Merritt & Bacon, 1997). Fig. 3 was generated using *ESPript* 3.0 (http://espript.ibcp.fr; Robert & Gouet, 2014). Superposition of GST molecular models was carried out using the method of Ferro & Hermans (1977). Figs. 4–10 were generated using *MolScript* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997). *LIGPLOT* (Wallace *et al.*, 1995) was used to identify residues involved in ligand binding.

The final model of DmGSTE6 contains residues 2–221 for polypeptide chain *A*, 2–222 for chain *B*, 3–221 for chains *C*, *D* and *H*, 3–222 for chains *E* and *F*, and 4–222 for chain *G*. Eight glutathione (GSH) molecules are bound to the eight subunits and together with 730 water molecules complete the structural model. The final model of DmGSTE7 contains residues 3–223 for chain *A*, 1–223 for chain *B*, two glutathione sulfenic acid (GSF) molecules and 403 water molecules.

Refinement statistics for both structures are given in Table 1. Coordinates and structure factors for DmGSTE6 have been deposited in the PDB (PDB entry

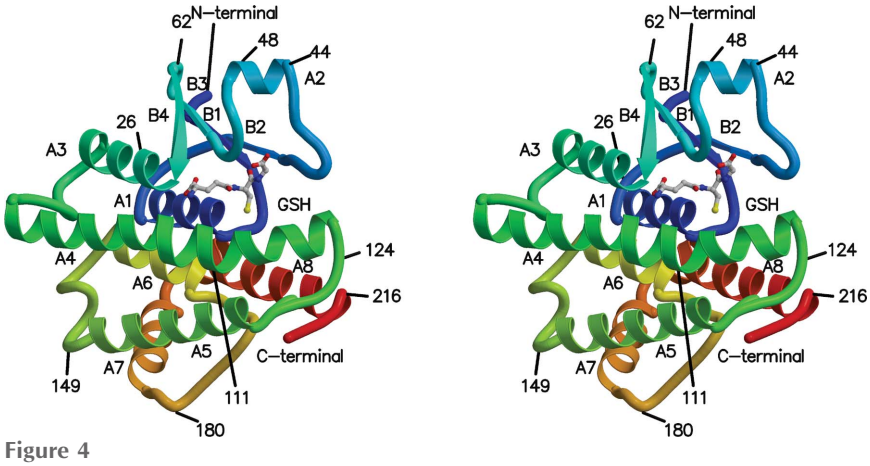


Figure 4
Stereoview of a DmGSTE6 monomer (chain *A*) showing the location of the secondary-structure elements as well as selected residues. The GSH ligand is rendered as sticks. The labels A and B denote α-helices and β-strands, respectively.

4pnf). Likewise, DmGSTE7 has been deposited with PDB code 4png.

2.4. Differential scanning calorimetry

Differential scanning calorimetry (DSC) measurements were performed on a MicroCal VP-Capillary DSC system with active cell volumes of 141 µl. All samples were exhaustively dialyzed against 100 m*M* KP_i pH 7.0 prior to DSC analysis. Calorimetric cells were kept under a 413 kPa nitrogen-gas excess pressure to prevent degassing during the scan. Protein concentrations were ∼10 µ*M* monomer in all experiments (based on ε_{280} = 29 300 and 28 020 M^{-1} cm^{-1} for the monomeric DmGSTE6 and DmGSTE7, respectively). The DSC thermograms were collected by scanning from 15 to 75°C at a rate of 1°C min^{-1}. Six buffer/buffer scans were performed before each experiment to assure proper equilibration of the calorimeter. The reversibility of the thermal transitions was evaluated by reheating the samples after fast cooling from the previous scan. The thermal transitions were found to be completely irreversible. The pre- and post-translational chemical baselines were derived and subtracted according to the procedure of Takahashi & Sturtevant (1981). *Origin* 7.0 (OriginLab) was used to normalize the data for protein concentration and for the subtraction of buffer reference scans.

3. Results and discussion

3.1. Molecular and crystal structures of DmGSTE6 and DmGSTE7

The initial models of DmGSTE6 contained several truncated side chains and lacked the residues near the C-terminus of this epsilon-class enzyme (see Fig. 3). Difference electron-density maps were used to correct those and other errors in the models of both DmGSTE6 and DmGSTE7. Side-chain atoms were not included unless positive difference density was observed for them. The models are frayed at the N- and the C-termini, where molecular motions presumably weaken the electron density.

The overall fold of both molecules is as expected for a GST, although the structures of these epsilon-class enzymes differ somewhat from those of other classes of GSTs (see below). The major elements of the DmGSTE6 subunit are shown in Fig. 4, where the strands and helices are labeled as in an early review (Dirr *et al.*, 1994). The eight subunits in the asymmetric unit form four GST dimers. A single dimer is found in the asymmetric unit of DmGSTE7. Thus, following the nomenclature used to describe GST dimers, the crystal structures described here contain copies of DmGSTE6-6 and DmGSTE7-7 homodimers. The twofold rotation axes relating the subunits in each dimer are noncrystallographic symmetry operations in these crystal forms.

The eight subunits of DmGSTE6 can be superposed with an r.m.s. difference from their average positions of 0.20 Å for the main-chain atoms of residues 15–206 (weak electron density for residues near the N- and C-termini is associated with greater structural variation or dynamics in these regions). The two subunits in the DmGSTE7 structure superpose with an r.m.s. difference of 0.28 Å for the main-chain atoms (residues 15–206). Because the crystallographically independent protein subunits are so similar, further discussion of the structures will focus on the *A* chains of each structure unless otherwise noted.

3.2. Comparison of the DmGSTE6 and DmGSTE7 ligand-binding sites

There are few sequence differences between the two enzymes around the G-site where GSH and GSF bind (see Fig. 5). Five residues form hydrogen bonds to GSH (and GSF). His53, Val55 (main-chain atoms) and Ser68 form such interactions with the ligand. In seven of the eight DmGSTE6 subunits Gln41 is hydrogen-bonded to GSH. A similar interaction is formed by Asn41 in DmGSTE7. In both cases, the side-chain amide is hydrogen-bonded to O31 of the ligand. Small conformational changes accommodate the additional methylene group in the glutamine side chain. Arg113 is positioned to form hydrogen bonds in all of the DmGSTE6 subunits and the *B* chain of DmGSTE7. A small change in the side-chain torsion angles places its guanidinium group too far away from the GSF in the *A* chain of DmGSTE7. In spite of this, overall, the hydrogen-bonding interactions between GSH (or GSF) and the proteins are very similar and are likely to contribute little to the functional differences between the DmGSTE6 and DmGSTE7 isozymes.

Identification of the residues involved in van der Waals contacts with ligands in the G-site is complicated by variation in the ten copies of the site in these two structures. Pro14, Thr54 and Phe108 are close (within 3.9 Å) to the ligands in all ten subunits, but Ile36 (Thr in DmGSTE7) is close to GSH or

GSF in four subunits and Ser12 and Asp67 are close in fewer subunits.

We have not been able to generate complexes of DmGSTE6 and DmGSTE7 with a GSH adduct. Without direct observation of the H-site, we have used previously determined structures to locate the likely H-site, the binding site for ligands conjugated to GSH. *S*-Hexylglutathione is the ligand in two epsilon- and delta-class structures: PDB entries 1pn9 and 2imk (see Table 2). There are four copies of the molecule in those two structures, with slightly different conformations of the alkyl chain. Superposition of the GSH portions of those molecules on the GSH or GSF in the ten subunits in DmGSTE6 and DmGSTE7 generates 40 possible protein structures and ligand combinations occupying what is likely to be the H-site for nonpolar ligands. Ile36, Phe108 and Leu120 are found in contact in a large number of the possible H-site models for DmGSTE6. In the other isozyme, DmGSTE7, Arg37, Phe120, Phe209 and Phe212 form the major portion of the H-site. Conformational differences between the two isozymes cause elimination of Phe108 and Ile36 from the H-site in DmGSTE7 and the addition of Arg37 and Phe212.

It should be emphasized that identification of all of these residues as contributors to the H-site is based on the assumption that the *S*-hexyl adduct is representative of the actual ligands for these isozymes. It could be that further conformational changes in the protein could produce a binding site with different residues and physical characteristics.

3.3. Comparison of the DmGSTE6 and DmGSTE7 dimer interfaces

In addition to isoform-dependent differences in their active sites, the dimer interface could, in principle, provide a pharmacological target for disruption of GST function. In particular, if the dimer interface were sufficiently different in epsilon GSTs compared with the human isoforms, this could provide additional therapeutic selectivity for antimalarial drugs or insecticides. For the canonical A-, P- or M-class GSTs the dimers are stable and not in a measurable equilibrium with monomers. Mutagenesis and structural studies have demonstrated the importance of a ball-in-socket interaction at the subunit interface of these isoforms (Stenberg *et al.*, 2000; Hegazy *et al.*, 2006). As in other insect GSTs (Wongsantichon & Ketterman, 2006), there is a lock-and-key structure that differs in its location when compared with other GST classes such as the alpha-class enzymes. In DmGSTE6 and DmGSTE7, the key residue is His101. The noncrystallographic twofold rotation axis relating the two subunits in the homodimer is nearby, and His101 from the other subunits forms part of the 'lock'. Other residues involved include Asp97 and Gln98 from the first subunit and Asp67, His69, Leu100 and Ser104 from the second.

3.4. Comparison of the structures of DmGSTE6 and DmGSTE7 with those of other epsilon-class GSTs

GSTs are classified largely on the basis of amino-acid sequence alignments (Mannervik *et al.*, 1985; Mashiyama *et al.*, 2014). Two unique classes, delta and epsilon, are found in insects (Ranson *et al.*, 2001, 2002; Ketterman *et al.*, 2011), and the crystal structures of several of these enzymes have been determined (for a selected set, see Table 2). What is unclear is which structural features correlate with the amino-acid sequence differences and whether these might be of functional significance. As a first step in addressing this issue, we have compared the three-dimensional structures of DmGSTE6 and

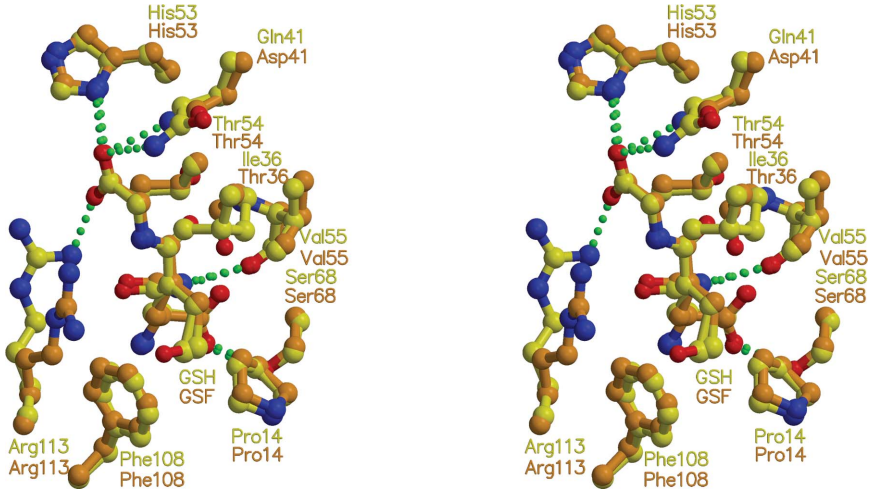


Figure 5
Stereoview of the ligands and residues in the G-sites of DmGSTE6 (yellow) and DmGSTE7 (orange). The GSH and GSF ligands were superposed to generate the figure.

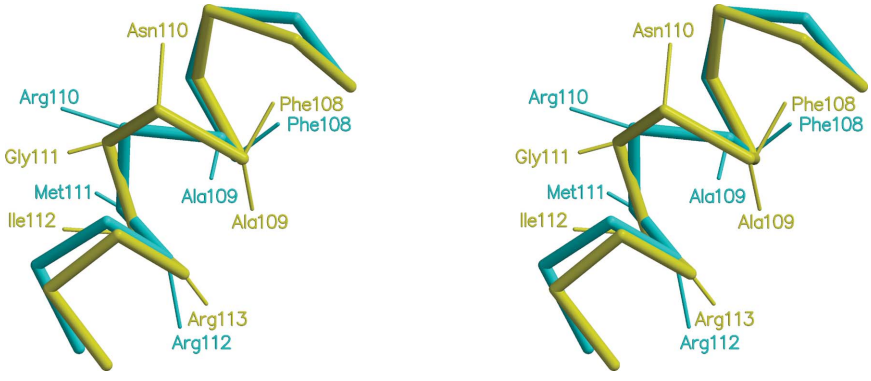


Figure 6
Stereoview C^{α} trace showing the effects of the additional residues near Gly111 in DmGSTE6 (yellow). The region in PDB entry 2il3 is shown in cyan.

DmGSTE7 with other GST structures. Only the *A* chains of the two structures were used for this analysis.

In addition to these structures from *Drosophila*, crystal structures of epsilon-class GSTs from housefly (*Musca domestica*; PDB entry 3vwx) and the mosquito *Anopheles gambiae* (PDB entries 2il3, 2imi and 2imk) are available for comparison (PDB entries 2imi and 2imk are complexes of the mosquito enzyme with glutathione and *S*-hexylglutathione). A crystal structure of a GST from *D. mojavensis* (PDB entry 4hi7) has also been referred to (Enzyme Function Initiative, unpublished work). Based on the amino-acid sequence of this unpublished structure, it is also an epsilon-class GST (unpublished work).

DmGSTE6 and the housefly enzyme are similar in that they both have residues at positions 111 and 149 (in the DmGSTE6 sequence) that can be viewed as insertions in the mosquito sequence. Residue 111 is in the middle of helix A4, and the additional residue causes one turn of the helix to expand into a turn with hydrogen bonding characteristic of a turn of π-helix (see Fig. 6). While Gly111 can be viewed as the additional residue in this turn, residues Ala109 and Asn110 are those most displaced from the equivalent positions in the mosquito protein. A complication in analyzing the conformation of this portion of the polypeptide is the fact that all three epsilon-class GSTs have an expanded turn of helix just before this insertion near residue 111. In the mosquito enzyme helix A4 is bent in this region, and in the other two structures the additional residue simply extends this region of expanded helix. The overall effect of the insertion on the conformation of helix A4 is not large since the polypeptide chains superpose fairly well at the residues preceding and following the insertion in the sequence.

At the other site where a residue is inserted, residue 149, the structural effects are ameliorated by this being located at the end of helix A5, and slight adjustments of the preceding and following residues cause only a local change in the course of the polypeptide chain (see Fig. 7). The *ClustalW* sequence alignment (Fig. 3) would place the misalignment of the amino-acid sequences at position 149 in the DmGSTE6 sequence, but the largest structural difference occurs at Gly148. Leu146 and Asp150 align well with their counterparts in the mosquito sequence, while Lys147 and Gln149 are somewhat displaced from Val146 and Asp147 in the mosquito structure. No residue in the mosquito molecule is close enough to be matched with Gly148.

In the mosquito enzymes with and without bound GSH, the A8 helix extends to within one or two residues of the C-terminus. In the *D. melanogaster* and housefly epsilon-class proteins, the last five residues do not take on a helical conformation. Instead, they take on an extended conformation where they interact with residues 125–128 in the turn between helices A4 and A5. These different conformations for the C-terminal residues appear to be species-related and not ligand-related.

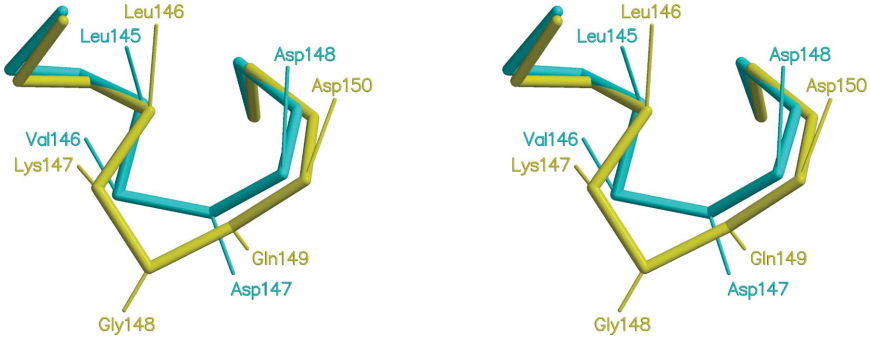


Figure 7
Stereoview C$^{\alpha}$ trace showing the effects of the additional residues near Gln149 in DmGSTE6 (yellow). The region in PDB entry 2il3 is shown in cyan.

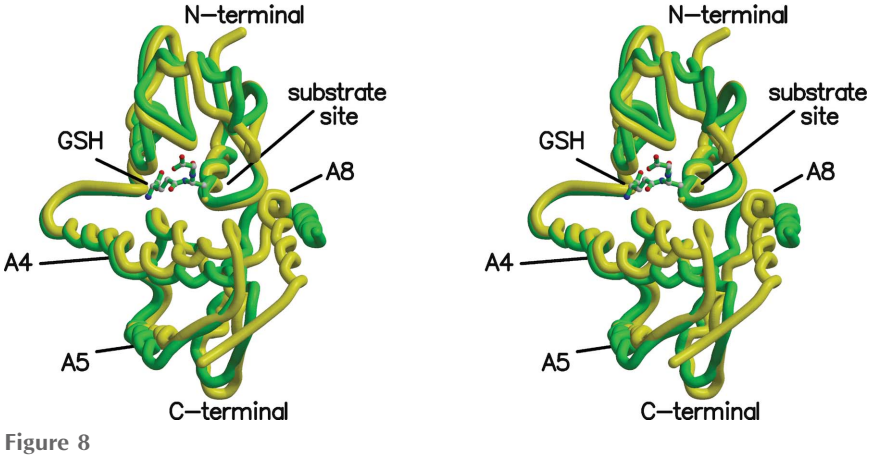


Figure 8
Stereoview showing the differences in positions of the A4 and A8 helices in epsilon- and delta-class GSTs. The epsilon-class GST (DmGSTE6) is shown in yellow amd the delta-class GST (PDB entry 1v2a) is shown in green.

3.5. Comparison of DmGSTE6 and DmGSTE7 with delta-class GSTs

In addition to a lack of information about functional differences among epsilon-class GSTs, those between epsilon-class, delta-class and other classes of GSTs are not well characterized. Comparison of the structures of the different classes of GSTs might suggest binding and active sites that could aid in understanding the biological roles of DmGSTE6 and DmGSTE7.

The first structural comparison of delta- and epsilon-class GSTs came with the structure analysis of AgGSTE2 from the mosquito *A. gambiae* (Wang *et al.*, 2008). Comparison of AgGSTE2 with AgGSTD1–6 showed that in the epsilon-class enzyme the A8 helix is longer and is slightly displaced from its position in the delta-class structure. This is also true for both DmGSTE6 and DmGSTE7 when compared with delta-class GSTs (see Fig. 8). Sequence comparisons (Fig. 3) show that the delta-

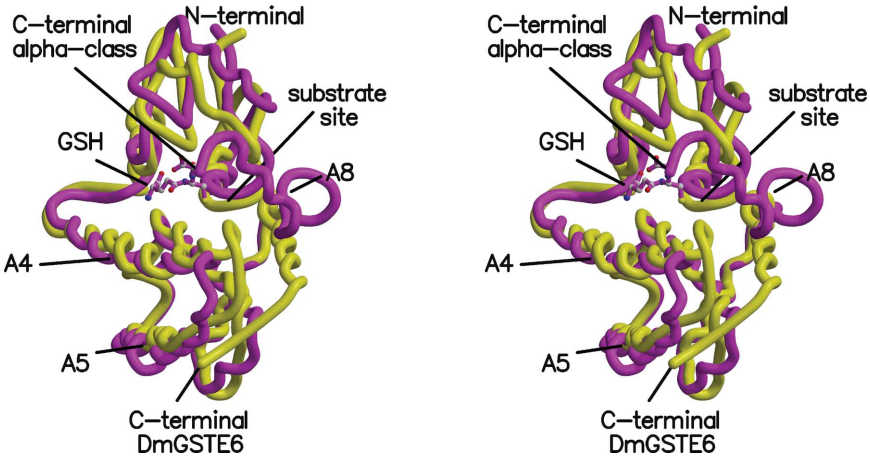


Figure 9
Stereoview showing the differences at the C-termini in epsilon-class and alpha-class GSTs. The epsilon-class GST (DmGSTE6) is shown in yellow and the alpha-class GST (PDB entry 1k3y) is shown in magenta.

class enzymes are three to nine residues shorter at their C-termini relative to the epsilon-class enzymes. These differences in the A8 helix create a slightly larger potential entrance to the active site for substrate molecules.

The C-terminal half of helix A4 is also displaced somewhat, and this alters the nature of the substrate-binding site (the H-site) and makes it more hydrophobic in the epsilon-class protein. At the same time, as pointed out in an analysis of DmGSTD1 from *D. melanogaster* (Low *et al.*, 2010), helix A4 is closer to GSH in the G-site in the epsilon-class enzyme.

A structure analysis of an epsilon-class GST from housefly (*M. domestica*; Nakamura *et al.*, 2013) concluded that entry to the substrate-binding pockets was limited by the locations of the A4 and A8 helices in epsilon-class GSTs. This does not agree with the suggestion by Wang *et al.* (2008). In both DmGSTE6 and DmGSTE7 A8 moves down to open access to the G-site, but A4 moves in to block it. More detailed analysis of the dynamics of these structures is needed in order to better understand this issue.

Something that was not mentioned in previous analyses is that the epsilon-class GSTs also differ from the delta-class proteins at the turn between the A6 and A7 helices. This is associated with the shift of the A8 helix, which is close to this portion of the polypeptide chain.

Finally, the delta- and epsilon-class structures known to date differ at their C-termini. Epsilon-class enzymes contain several residues beyond the position where the A8 helix is terminated in delta-GSTs. In DmGSTE6, DmGSTE7 and the housefly enzyme, these residues are extended towards the solvent, pack against the surface of the protein and interact with residues in the turn between A4 and A5 (see Fig. 8). In the mosquito enzymes, these residues take on helical conformations and lengthen the A8 helix. In alpha-class GSTs complexed with GSH (see Fig. 9), the terminal residues form a helix that closes off the substrate-binding site. The observed differences between the C-terminal residues in delta- and epsilon-class GSTs in the absence of substrate could play a role in the different functions of the isozymes

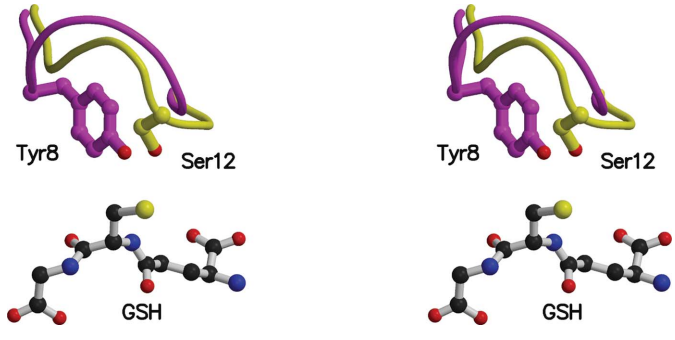


Figure 10
Stereoview showing the overlap of Ser12 in epsilon-class GSTs (DmGSTE6, yellow) and Tyr8 in alpha-class GSTs (PDB entry 1k3y, magenta).

3.6. Structural correlates of sequence-based phylogenies

The assignment of GST molecules to the various classes is based on sequence alignments (Mannervik *et al.*, 1985; Mashiyama *et al.*, 2014). A more complete description of the similarities and dissimilarities of the GST classes can be obtained by combining three-dimensional structure alignments with sequence and functional information. Kosloff *et al.* (2006) identified structural features and motifs that permit the assignment of the GST class on the basis of the three-dimensional molecular structures. Their analysis included the only delta-class enzyme for which a structure was available at the time. Subsequently, structures for several additional delta-class and epsilon-class GSTs have been determined. The following paragraphs indicate how the table presented by Kosloff *et al.* (2006) can be updated to include features of the delta-class and epsilon-class GSTs.

The first structural point that differentiates the GST classes is the identity of the hydroxyl group that hydrogen bonds to GSH and is important for catalysis (Armstrong, 1997). In many GSTs tyrosine fulfills this role, but in delta-class and epsilon-class GSTs (along with others) serine provides the nucleophile. This is only surmised in many of the enzymes, but mutagenesis experiments deleting the serine have reduced the activity of delta-class and epsilon-class GSTs from silkworm (Yamamoto *et al.*, 2013).

These catalytic serines are 4–5 residues away from tyrosines that are often aligned with the catalytic tyrosines in sequence comparisons. What is most important for the enzymatic function is that the hydroxyl O atoms of the catalytic tyrosines and serines superpose in three-dimensional space. The catalytic serines and tyrosines are on opposite ends of a loop in the polypeptide chain, and the longer side chains of the tyrosines

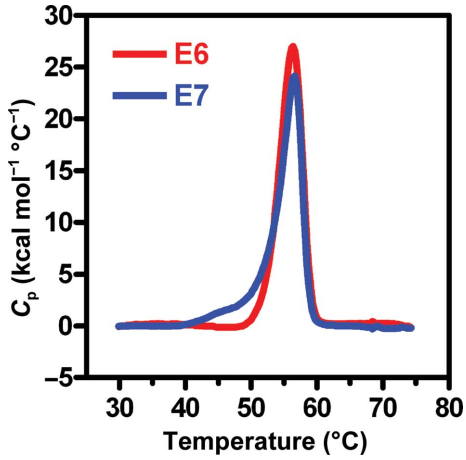


Figure 11
Differential scanning calorimetry of DmGSTE6 and DmGSTE7. The T_m for DmGSTEE6 is 56.4°C and that for DmGSTEE7 is 56.6°C. The thermal stability does not differ for these isoforms.

allow them to reach to the serine hydroxyl position (see Fig. 10; Atkinson & Babbitt, 2009).

A second characteristic structural motif of the GSTs is the presence or absence of a loop of four residues between B2 and A2 seen in mu-class GSTs. The delta and epsilon classes do not have such a loop. Also, as stated before, the residues forming the lock-and-key motif in the dimer interface are not in the same location as in the mammalian GSTs (see above).

A sequence motif of some importance in GST classification schemes is in helix A3, where the sequences SNAIL or TRAIL are found. In delta GSTs, this sequence is S_{65}RAIC, as it is in phi-class GSTs. The sequence S_{68}HAII is unique so far for epsilon-class enzymes, although it is very similar to the SFAII sequence found in zeta-class GSTs. When the three-dimensional structures are compared, the significance of this sequence motif becomes uncertain because the residues in the delta and epsilon classes do not superpose well with the motif in other classes. It should be noted that the residues in this motif also form part of the G-site and part of the lock-and-key motif. Thus, they provide a linkage between the dimer interface and the G-site.

In many GSTs, an acidic side chain in the A4 helix has been implicated in GSH binding. In the delta and epsilon classes there is no carboxylate near the active site positioned to carry out the function of this group in the other GSTs. In *ClustalW* sequence alignments using default weights, a glutamic acid (residue 103 in the DmGSTE6 sequence) is aligned with an aspartic acid in the delta-class sequences and the aspartic acids believed to be catalytically important in the alpha class and other GSTs (see Fig. 3). Based on sequence alignment, Glu103 would be considered to be a catalytic residue, but a structural comparison shows that this is unlikely to be the case. This is also true for the aspartic acid in the delta-class enzymes. In these classes, Glu103 (and the equivalent aspartic acid in the delta-class molecules) is positioned on the A4 helix, pointing away from the active site. A major structural reorganization would be necessary to place these carboxylates near the GSH-binding site.

3.7. Thermal stability

A sensitive probe of structural differences among proteins is thermal stability, as measured by differential scanning calorimetry (DSC). Subtle differences in structure can lead to differences in the melting temperatures among proteins, owing to differences in the ΔH for unfolding. In order to examine this possibility, DSC was performed with the ligand-free enzymes (Fig. 11). The melting temperatures, T_m, for DmGSTE6 and DmGSTE7 are essentially identical at 56.4 and 56.6°C, respectively, indicating that their thermodynamic stabilities parallel their structural similarity. For comparison, the T_m for GSTA1-1 is 58.6°C and that for GSTA4-4 is 51.2°C, indicating that even GSTs within a class can have significantly different thermal stabilities (Honaker *et al.*, 2013).

4. Conclusions

Despite the modest differences in substrate selectivities reported for these isoforms on the basis of a limited range of substrates (Kjellander *et al.*, 2014), the structures of DmGSTE6 and DmGSTE7 are nearly invariant and they are also very similar to the few other epsilon-class GSTs for which structures are available. Although it is expected that the GSH-binding sites would be structurally conserved in other GST classes, the substrate-binding sites (H-sites) might be expected to vary more than is observed here between these isoforms. In addition, the structure of the dimer interface in each of these epsilon GSTs is similar to other epsilon-class GSTs and the delta-class GSTs, but distinctly different from the canonical mammalian GSTs. The thermodynamic stability of the two epsilon GSTs is essentially identical, further emphasizing their structural similarity. If these two epsilon-class isoforms have different physiological roles that involve endogenous substrates then the substrate-free structures do not suggest obvious sources of substrate specificity, and induced fit may be required to achieve such specificity. Alternatively, if these isoforms are promiscuous detoxication enzymes then the structures suggest that they would have highly overlapping selectivities or they undergo very different conformational changes in response to different substrates. The results provide a striking example where functional diversity is not necessarily predictable on the basis of structure. Alternatively, these two isoforms might be functionally as well as structurally redundant. Additional work is required to understand the physiological functions of these enzymes.

Acknowledgements

The research reported in this publication was supported by the Department of Medicinal Chemistry and by The Swedish Research Council. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of

Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. Differential scanning calorimetry was performed in the Analytical Biopharmacy Core Facility, University of Washington, which is part of the Center for Intracellular Delivery of Biologics.

References

Armstrong, R. N. (1997). *Chem. Res. Toxicol.* **10**, 2–18.
Atkinson, H. J. & Babbitt, P. C. (2009). *Biochemistry*, **48**, 11108–11116.
Balogh, L. M., Le Trong, I., Kripps, K. A., Tars, K., Stenkamp, R. E., Mannervik, R. & Atkins, W. M. (2009). *Biochemistry*, **48**, 7698–7704.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chen, L., Hall, P. R., Zhou, X. E., Ranson, H., Hemingway, J. & Meehan, E. J. (2003). *Acta Cryst.* D**59**, 2211–2217.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Deponte, M. & Becker, K. (2005). *Methods Enzymol.* **401**, 241–253.
Dirr, H., Reinemer, P. & Huber, R. (1994). *Eur. J. Biochem.* **220**, 645–661.
Dudas, S. P. & Arking, R. (1995). *J. Gerontol. A Biol. Sci. Med. Sci.* **50**, B117–B127.
Ferro, D. R. & Hermans, J. (1977). *Acta Cryst.* A**33**, 345–347.
Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010). *Nucleic Acids Res.* **38**, W695–W699.
Harwaldt, P., Rahlfs, S. & Becker, K. (2002). *Biol. Chem.* **383**, 821–830.
Hayes, J. D., Flanagan, J. U. & Jowsey, I. R. (2005). *Annu. Rev. Pharmacol. Toxicol.* **45**, 51–88.
Hegazy, U. M., Hellman, U. & Mannervik, B. (2006). *Chem. Biol.* **13**, 929–936.
Henderson, C. J. & Wolf, C. R. (2005). *Methods Enzymol.* **401**, 115–135.
Honaker, M. T., Acchione, M., Zhang, W., Mannervik, B. & Atkins, W. M. (2013). *J. Biol. Chem.* **288**, 18599–18611.
Johnson, W. W., Liu, S., Ji, X., Gilliland, G. L. & Armstrong, R. N. (1993). *J. Biol. Chem.* **268**, 11508–11511.
Jones, C. M., Toé, H. K., Sanou, A., Namountougou, M., Hughes, A., Diabaté, A., Dabiré, R., Simard, F. & Ranson, H. (2012). *PLoS One*, **7**, e45995.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Ketterman, A. J., Saisawang, C. & Wongsantichon, J. (2011). *Drug Metab. Rev.* **43**, 253–265.
Kjellander, M., Mazari, A. M., Boman, M., Mannervik, B. & Johansson, G. (2014). *Anal. Biochem.* **446**, 59–63.
Kosloff, M. *et al.* (2006). *Proteins*, **65**, 527–537.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Landis, G., Shen, J. & Tower, J. (2012). *Aging (Albany NY)*, **4**, 768–789.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Le Trong, I., Stenkamp, R. E., Ibarra, C., Atkins, W. M. & Adman, E. T. (2002). *Proteins*, **48**, 618–627.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Low, W. Y., Feil, S. C., Ng, H. L., Gorman, M. A., Morton, C. J., Pyke, J., McConville, M. J., Bieri, M., Mok, Y.-F., Robin, C., Gooley, P. R., Parker, M. W. & Batterham, P. (2010). *J. Mol. Biol.* **399**, 358–366.
Mahajan, S. & Atkins, W. M. (2005). *Cell. Mol. Life Sci.* **62**, 1221–1233.
Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. & Jörnvall, H. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 7202–7206.
Mashiyama, S. T. *et al.* (2014). *PLoS Biol.* **12**, e1001843.
McElwee, J. J., Schuster, E., Blanc, E., Piper, M. D., Thomas, J. H., Patel, D. S., Selman, C., Withers, D. J., Thornton, J. M., Partridge, L. & Gems, D. (2007). *Genome Biol.* **8**, R132.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Mohring, F., Pretzel, J., Jortzik, E. & Becker, K. (2014). *Curr. Med. Chem.* **21**, 1728–1756.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nakamura, C., Yajima, S., Miyamoto, T. & Sue, M. (2013). *Biochem. Biophys. Res. Commun.* **430**, 1206–1211.
Norrgård, M. A., Ivarsson, Y., Tars, K. & Mannervik, B. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4876–4881.
Oakley, A. J., Harnnoi, T., Udomsinprasert, R., Jirajaroenrat, K., Ketterman, A. J. & Wilce, M. C. J. (2001). *Protein Sci.* **10**, 2176–2185.
Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., Sharakhova, M. V., Unger, M. F., Collins, F. H. & Feyereisen, R. (2002). *Science*, **298**, 179–181.
Ranson, H., Rossiter, L., Ortelli, F., Jensen, B., Wang, X., Roth, C. W., Collins, F. H. & Hemingway, J. (2001). *Biochem. J.* **359**, 295–304.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Reddy, B. P., Prasad, G. B. & Raghavendra, K. (2011). *Comput. Biol. Chem.* **35**, 114–120.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Saisawang, C., Wongsantichon, J. & Ketterman, A. J. (2012). *Biochem. J.* **442**, 181–190.
Sharma, R., Yang, Y., Sharma, A., Awasthi, S. & Awasthi, Y. C. (2004). *Antioxid. Redox Signal.* **6**, 289–300.
Stenberg, G., Abdalla, A.-M. & Mannervik, B. (2000). *Biochem. Biophys. Res. Commun.* **271**, 59–63.
Takahashi, K. & Sturtevant, J. M. (1981). *Biochemistry*, **20**, 6185–6190.
Tars, K., Larsson, A. K., Shokeer, A., Olin, B., Mannervik, B. & Kleywegt, G. J. (2006). *J. Mol. Biol.* **355**, 96–105.
Tars, K., Olin, B. & Mannervik, B. (2010). *J. Mol. Biol.* **397**, 332–340.
Udomsinprasert, R., Pongjaroenkit, S., Wongsantichon, J., Oakley, A. J., Prapanthadara, L.-A., Wilce, M. C. & Ketterman, A. J. (2005). *Biochem. J.* **388**, 763–771.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng. Des. Sel.* **8**, 127–134.
Wang, Y., Qiu, L., Ranson, H., Lumjuan, N., Hemingway, J., Setzer, W. N., Meehan, E. J. & Chen, L. (2008). *J. Struct. Biol.* **164**, 228–235.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wongsantichon, J. & Ketterman, A. J. (2006). *Biochem. J.* **394**, 135–144.
Wongsantichon, J., Robinson, R. C. & Ketterman, A. J. (2012). *Arch. Biochem. Biophys.* **521**, 77–83.
Yamamoto, K., Aso, Y. & Yamada, N. (2013). *Insect Mol. Biol.* **22**, 523–531.
Zhang, J. H., Grek, C., Ye, Z.-W., Manevich, Y., Tew, K. D. & Townsend, D. M. (2014). *Adv. Cancer Res.* **122**, 143–175.