The EMBO Journal vol.14 no.10 pp.2133–2143, 1995

Crystal structure of a theta-class glutathione transferase

Matthew C.J.Wilce, Philip G.Board[1], Susanne C.Feil and Michael W.Parker[2]

The Ian Potter Foundation Protein Crystallography Laboratory, St Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy, Victoria 3065 and [1]John Curtin School of Medical Research, Australian National University, Canberra, ACT 2601, Australia

[2]Corresponding author

Communicated by W.Kühlbrandt

Glutathione S-transferases (GSTs) are a family of enzymes involved in the cellular detoxification of xenotoxins. Cytosolic GSTs have been grouped into four evolutionary classes for which there are representative crystal structures of three of them. Here we report the first crystal structure of a theta-class GST. So far, all available GST crystal structures suggest that a strictly conserved tyrosine near the N-terminus plays a critical role in the reaction mechanism and such a role has been convincingly demonstrated by site-directed mutagenesis. Surprisingly, the equivalent residue in the theta-class structure is not in the active site, but its role appears to have been replaced by either a nearby serine or by another tyrosine residue located in the C-terminal domain of the enzyme.

Key words: detoxification/evolution/glutathione S-transferase/structural similarity/X-ray crystallography

Introduction

Glutathione S-transferases (GSTs; EC 2.5.1.18) are a ubiquitous family of multi-functional enzymes that conjugate electrophilic substrates to the tripeptide glutathione (GSH) (reviewed by Mannervik and Danielson, 1988). Their substrates include a wide number of exogenous and endogenous hydrophobic electrophiles. The conjugation increases the solubility of the target molecule, thus facilitating the excretion of the molecule from the organism. Mammalian cytosolic GSTs exist either as homo- or heterodimers with a subunit mol. wt of ~25 kDa. They can be classified into four distinct families, alpha, mu, pi and theta, based on studies of substrate specificity and primary structures. The amino acid sequence identity between any two members within a class is typically >70%, whereas the figure is typically <30% between classes. Theta-class isoenzymes have been identified in several non-mammalian species and a fifth class (sigma) has been proposed in squid (Buetler and Eaton, 1992). GSTs have been implicated in the development of the resistance of cells and organisms towards drugs, insecticides, herbicides and antibiotics, and hence have been the subject of intense research over the last few years (for example, see Mannervik and Danielson, 1988; Board *et al.*, 1990; Fournier *et al.*, 1992; Wilce and Parker, 1994).

There are representative crystal structures for three of the four cytosolic mammalian GST classes. These include pig (Reinemer *et al.*, 1991), human (Reinemer *et al.*, 1992) and mouse (García-Sáez *et al.*, 1994) pi-class GSTs, rat (Ji *et al.*, 1992) and human (Raghunathan *et al.*, 1994) mu-class GSTs, and human alpha-class GST (Sinning *et al.*, 1993). The overall polypeptide fold is very similar between the crystal structures, but each class exhibits unique features, particularly about the active site and at the C-terminus (Wilce and Parker, 1994). Each subunit is characterized by two distinct domains and possesses an active site that acts independently of the other subunit (Mannervik and Danielson, 1988; Wilce and Parker, 1994). The N-terminal domain of ~80 residues adopts an α/β topology and contributes most of the contacts to GSH. The topology of this domain resembles that of some other enzymes that bind GSH (Reinemer *et al.*, 1991; Sinning *et al.*, 1993). The C-terminal domain is all α-helical and provides some of the contacts to the hydrophobic binding site which lies adjacent to the GSH binding site.

Even though considerable progress has been made in elucidating the nature of the catalytic mechanism, many of the fundamental details remain poorly understood. The central features of the catalytic mechanism are that the nucleophilic species in the active site is the thiolate anion of GSH and that the reaction is sequential with the addition of GSH to the electrophilic substrate occurring in the ternary enzyme complex (Armstrong, 1991). Mutagenesis studies have shown that a conserved tyrosine near the N-terminus is responsible for activation of GSH by promoting thiolate formation (Stenberg *et al.*, 1991; Kolm *et al.*, 1992; Kong *et al.*, 1992a,b; Liu *et al.*, 1992; Manoharan *et al.*, 1992a,b; Penington and Rule, 1992; Wang *et al.*, 1992). This residue is in close proximity to the sulfur atom of GSH in all the crystal structures (Reinemer *et al.*, 1991, 1992; Ji *et al.*, 1992; Sinning *et al.*, 1993; García-Sáez *et al.*, 1994; Raghunathan *et al.*, 1994). In the alpha-class enzyme, a conserved arginine residue at position 15 also contributes significantly to the stabilization of the GSH thiolate anion (Mannervik *et al.*, 1993).

GSTs have been identified in insects, although they are not as well characterized as their mammalian counterparts (Clark, 1989). Like the mammalian enzymes, they are dimers of comparable molecular weight. GSTs from *Drosophila*, house flies and sheep blowflies share >80% pairwise sequence identity, but share <20% sequence identity with the alpha, mu and pi mammalian enzymes (Fournier *et al.*, 1992; Wilce and Parker, 1994). On the basis of sequence comparisons, insect GSTs have been classified as belonging to the theta-class family (Hiratsuka *et al.*, 1990; Meyer *et al.*, 1991; Pemble and Taylor, 1992). The theta-class GSTs have been proposed as the

Table I. Sequence identities between the N-terminal domains of GSTs

	BF	HF	FF	BuF	HT	RT	FI	PL	BA	HP	RM	HA
Blowfly (BF)	100	99	93	51	41	39	30	40	28	18	21	20
Housefly (HF)		100	84	49	41	38	31	38	28	18	21	20
Fruitfly (FF)			100	46	37	37	29	36	25	16	18	18
Butterfly (BuF)				100	43	38	32	38	24	12	22	13
Human theta-class (HT)					100	89	28	39	29	17	22	16
Rat theta-class (RT)						100	24	36	24	18	18	18
Fish (FI)							100	36	25	18	16	12
Plant (PL)								100	28	16	25	20
Bacteria (BA)									100	16	18	15
Human pi-class (HP)										100	28	25
Rat mu-class (RM)											100	26
Human alpha-class (HA)												100

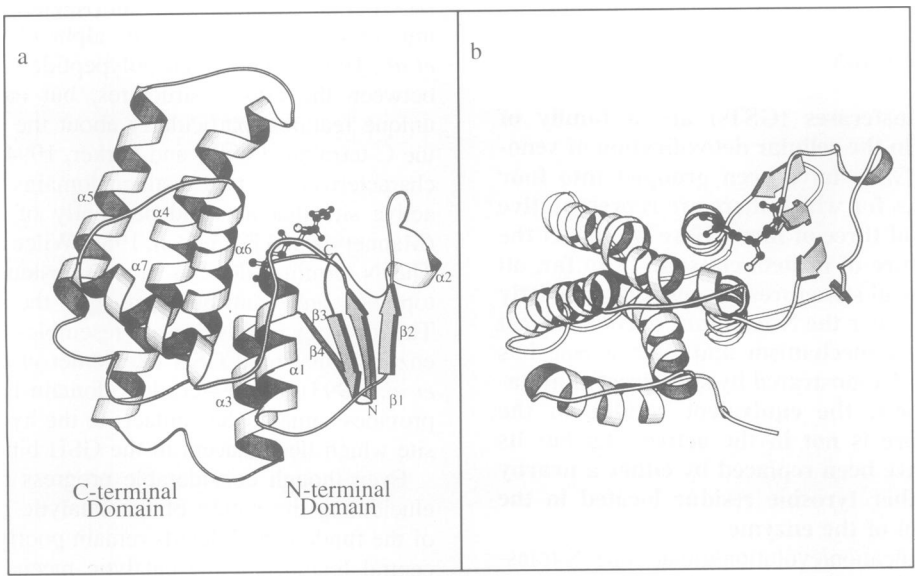

Fig. 1. Ribbon representation of the *L.cuprina* GST monomer. Residues were assigned to secondary structural elements according to criteria defined by Kabsch and Sander (1983). The locations of α-helices (denoted α) and β-strands (denoted β) are indicated. The bound GSH is shown in ball-and-stick fashion. Two orthogonal views are shown. The figures were produced by the computer program MOLSCRIPT (Kraulis, 1991).

evolutionary forerunner of the alpha, mu and pi enzymes, based on the apparent distribution of the former in a diverse range of organisms including bacteria, yeast, plants and insects (Buetler and Eaton, 1992; Pemble and Taylor, 1992). To date, there have been no crystal structures presented of either a theta-class GST or a non-mammalian GST. Here we present the crystal structure of a theta-class GST from *Lucilia cuprina*, the Australian sheep blowfly.

Results and discussion

Protein structure

The crystal structure of *Lucilia* GST adopts the canonical GST fold, despite its low sequence identity with other GSTs for which there are crystal structures available (see Table I). The structure of the monomer can be divided into two domains separated by a short linker region of six residues (Figure 1). The smaller N-terminal domain (residues 1–78) consists of a central four-stranded β-sheet flanked on one side by two α-helices and, on the other side, a bent irregular helical structure. The mixed β-sheet adopts a – 1 + 2 + 1 topology. There is a *cis*-proline bend at residue 53; the equivalent residue in all GST crystal structures adopts the *cis* conformation (Reinemer *et al.*, 1991, 1992; Ji *et al.*, 1992; Sinning *et al.*, 1993; García-Sáez *et al.*, 1994; Raghunathan *et al.*, 1994). This residue appears to be critical for ensuring main-chain hydrogen bonding about this region to the GSH substrate. The larger C-terminal domain (residues 85–208) consists of a right-handed bundle of four α-helices. Helix 4 is slightly bent due to the insertion of a glycine residue at position 102. A large proportion of helix 6 is buried within the core of the domain and consists predominantly of hydrophobic side chains. The interface between the two domains is characterized by a mixture of salt links, hydrogen bonds, water-mediated contacts and van der Waals contacts. Within the linker region, the aromatic side chain of Phe 83 is wedged between the two domains, reminiscent of the positioning of the equivalent residue in the pi- (Tyr 79) and alpha-class GSTs (Tyr 81). There are two buried charged residues (Asp 100 and Asp 156), both of which are involved in extensive hydrogen bonding and salt links to other residues and buried water molecules.

Comparison with other GSTs

Lucilia GST shares a high degree of sequence identity with other insect GSTs, whereas the degree of sequence identity with the mammalian alpha-, mu- and pi-class

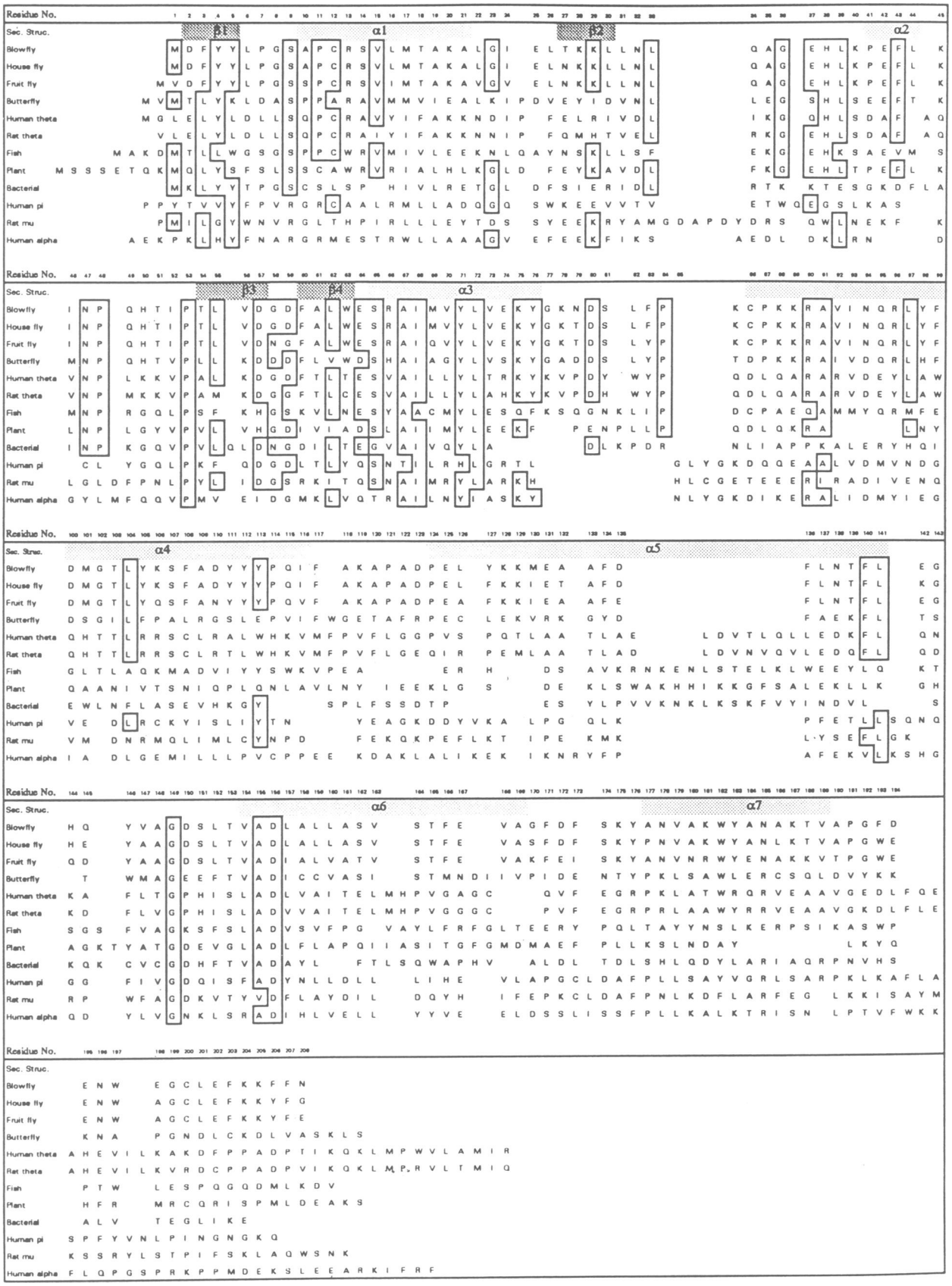

Fig. 2. Sequence alignment of GSTs showing the secondary structure elements of the *Lucilia* enzyme. GSTs for which there are available crystal structures were aligned based on three-dimensional structural superpositions; otherwise, alignments were performed using the PILEUP program in the GCG package [Genetics Computer Group Program Manual for the GCG Package, Version 7 (575 Science Drive, Madison, WI, 1991)]. Sequences were extracted from the SWISSPROT data base (Bairoch and Boeckmann, 1991). The sequence numbering refers to the *Lucilia* amino acid sequence. Residues which are conserved in at least seven of the sequences listed have been boxed.

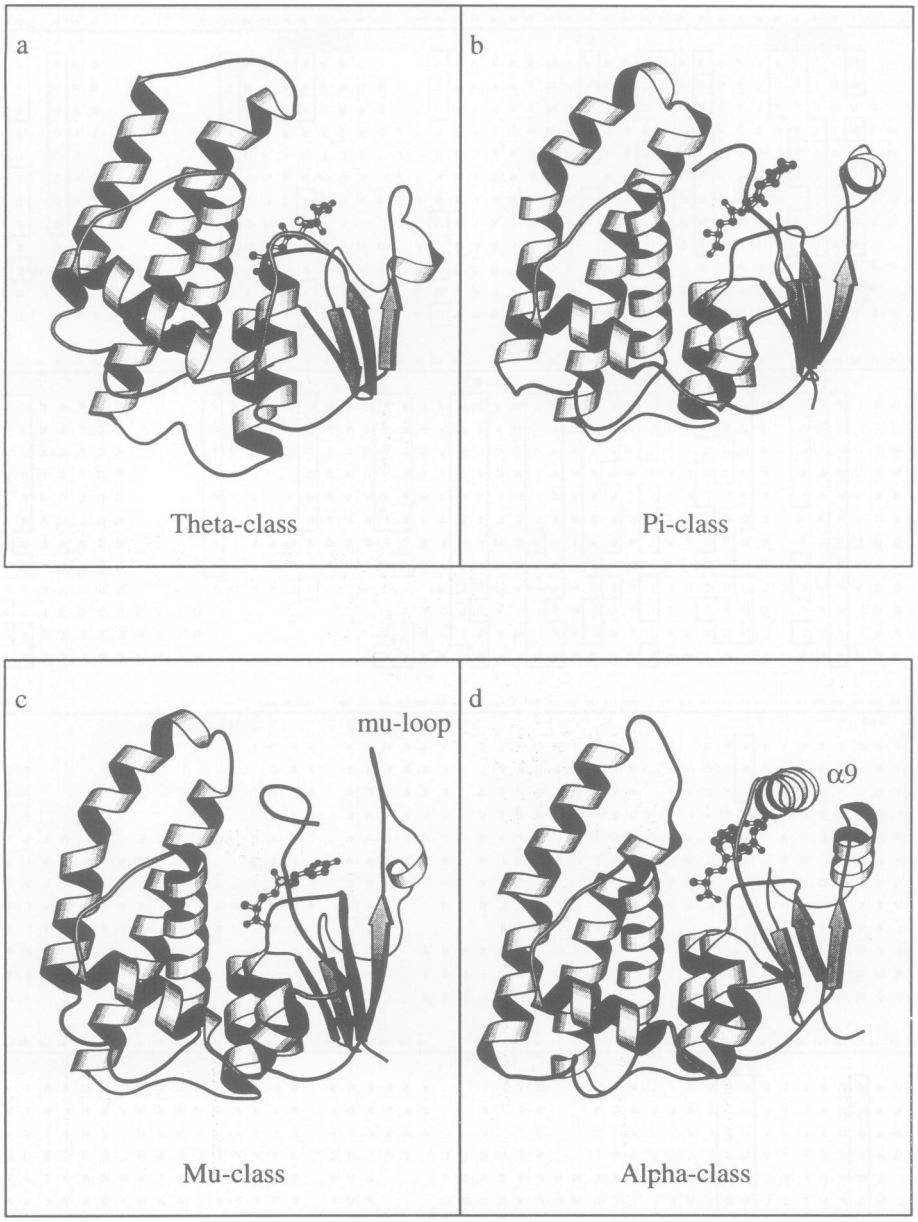

Fig. 3. Ribbon representation of GST crystal structures. Residues were assigned to secondary structural elements according to criteria defined by Kabsch and Sander (1983). The inhibitors are shown in ball-and-stick fashion. The figures were produced by the computer program MOLSCRIPT (Kraulis, 1991). (**a**) *Lucilia cuprina* GST. (**b**) Human pi-class GST (Reinemer *et al.*, 1992). (**c**) Rat mu-class GST (Ji *et al.*, 1992). (**d**) Human alpha-class GST (Sinning *et al.*, 1993).

GSTs is very low (Figure 2, Table I). The sequence identity between the insect GSTs and the mammalian theta-class enzymes is >37% in the N-terminal domain region (Table I). However, there is <20% sequence identity in the C-terminal domain comparisons (data not shown). Similar low levels of pairwise sequence identities are found in inter-class comparisons of GST C-terminal domains due presumably to the fact that the majority of active site residues are located in the N-terminal domain rather than the C-terminal domain, and the latter domain is largely responsible for the observed intra-class substrate diversity characteristic of GSTs. Phylogenetic analyses of GST sequences by different groups suggest that either the mu- (Pemble and Taylor, 1992) or pi-class (Buetler and Eaton, 1992) enzyme was first to diverge from the ancestral theta-class enzyme. However, our own unpublished analysis (V.Ross, S.Easteal and P.Board) suggests that the time of divergence of the alpha, mu, pi and sigma classes from the theta class is so ancient that the order of their divergence cannot be reliably assessed with the available data. It is not obvious from a close comparison of the crystal structures which class the theta fold most resembles (see Figures 2 and 3, Table II).

There are a number of obvious differences between the *Lucilia* GST crystal structure and the mammalian GST crystal structures (Figure 3). The *Lucilia* structure does not have the extended mu-loop structure characteristic of the mu-class enzyme or the C-terminal helix characteristic of the alpha-class enzyme. Unlike the mammalian enzymes, helix 5 is shorter and is not bent. The conformation of the linker region between the two domains, the loop connecting helices 4 and 5, and the absence of helix

Table II. Comparison of GST crystal structures

R.m.s. deviations (Å) between protein units

		alpha	pi	mu
pi	dimer	2.07 (361/404)		
	monomer	1.63 (175/202)		
	N-terminal domain	1.09 (68/75)		
	C-terminal domain	1.34 (101/127)		
mu	dimer	2.11 (375/420)	2.01 (349/404)	
	monomer	1.91 (174/210)	1.37 (178/202)	
	N-terminal domain	1.17 (67/82)	1.13 (59/75)	
	C-terminal domain	1.55 (98/128)	1.26 (115/127)	
theta	dimer	2.65 (220/380)	2.26 (289/380)	2.09 (223/380)
	monomer	2.31 (131/190)	1.87 (132/190)	1.91 (149/190)
	N-terminal domain	1.43 (68/75)	1.43 (60/75)	1.43 (61/75)
	C-terminal domain	1.35 (80/115)	1.12 (75/115)	1.33 (77/115)

Rotation between structural units

		alpha	pi	mu
pi	dimer	4.60		
	N-terminal domain	8.47		
	C-terminal domain	2.29		
mu	dimer	3.03	6.28	
	N-terminal domain	8.09	4.62	
	C-terminal domain	2.12	1.71	
theta	dimer	9.28	7.57	9.17
	N-terminal domain	19.52	15.45	10.55
	C-terminal domain	5.79	2.81	6.74

The comparisons were made using the α-carbon atoms of whole dimers, monomers and domains. The number of equivalent atoms used to make each comparison and the total number of atoms available are shown in parentheses. For the rotations, pairs of monomers were superimposed and then the angle required to optimally fit the units was determined. All calculations were performed using LSQMAN from the O-suite of programs (Jones *et al.*, 1991).

8 in the *Lucilia* structure are other obvious differences. The results of superpositions of the *Lucilia* GST crystal structure against the mammalian GST structures are shown in Table II. Each pair of monomers can be aligned with a root-mean-square (r.m.s.) deviation in α-carbon positions of between 1.4 and 2.3 Å. The theta-class enzyme shows the largest r.m.s. deviations, particularly in the C-terminal domain where <80 atoms could be matched due to differences in helix lengths and changes to helix curvatures. The domain rotations between the theta and other classes are very large, reflecting the fact that the V-shaped interface between the two domains seen in the mammalian enzymes is replaced by a much more parallel interface in the *Lucilia* structure.

Recently, Dirr *et al.* (1994) have compiled a list of 26 residues that are invariant amongst the known GST crystal structures. Only six of these residues are present in the *Lucilia* structure. These include Tyr 5, Pro 53, Ile 68, Leu 141, Gly 149 and Asp 156. With the exception of Tyr 5, all the other residues have been assigned purely structural roles. Of particular note is Ala 10, which is a glycine in the mammalian GSTs. This glycine adopts a conformation that would be energetically unfavourable for non-glycine residues, and is thought to do so because the carbonyl oxygen would otherwise point towards the active site and hydrogen bond to the catalytically important tyrosine residue near the N-terminus (Sinning *et al.*, 1993). In *Lucilia* GST, the carbonyl of the alanine points away from the active site, but nevertheless manages to adopt an energetically favourable conformation.

In the sequence alignments of Figure 2, there are two major insertions of sequence in the mammalian theta-class enzymes relative to the *Lucilia* enzyme. There is an eight residue insertion in the helix 5 region which is likely to extend the length of the helix in the mammalian enzymes, as judged by secondary structure prediction algorithms (data not shown). Secondary structure prediction algorithms suggest that the extended C-termini of the mammalian theta-class enzymes might adopt an amphipathic helical conformation reminiscent of the situation in the alpha-class enzyme (Sinning *et al.*, 1993; Figure 3d).

Subunit contacts

The functional dimer is related by exact crystallographic symmetry. The molecule is globular with approximate dimensions of 55×60×50 Å. The dimer interface is fairly extensive with a total of ~2870 Å^2 of surface area buried between monomers. Similar values are obtained with the mammalian GSTs (Wilce and Parker, 1994). The side-on view of the dimer, shown in Figure 4, shows the interface to be roughly V-shaped, and it is formed by a mixture of hydrophilic and hydrophobic interactions. At first sight, the monomer–monomer packing observed in the *Lucilia* dimer resembles that found in the other GST structures. However, the detailed interactions are quite different. Unlike the other GST structures, there are few interactions at the periphery at the interface, whereas in the middle, close to the 2-fold axis, there is a high density of interactions. There are surprisingly very few direct polar interactions between side chains. There are six residues involved in hydrogen-bonding interactions (Gln 49 to Thr 103, Trp 63 to Gln 95, Tyr 71 to Lys 88) and one salt

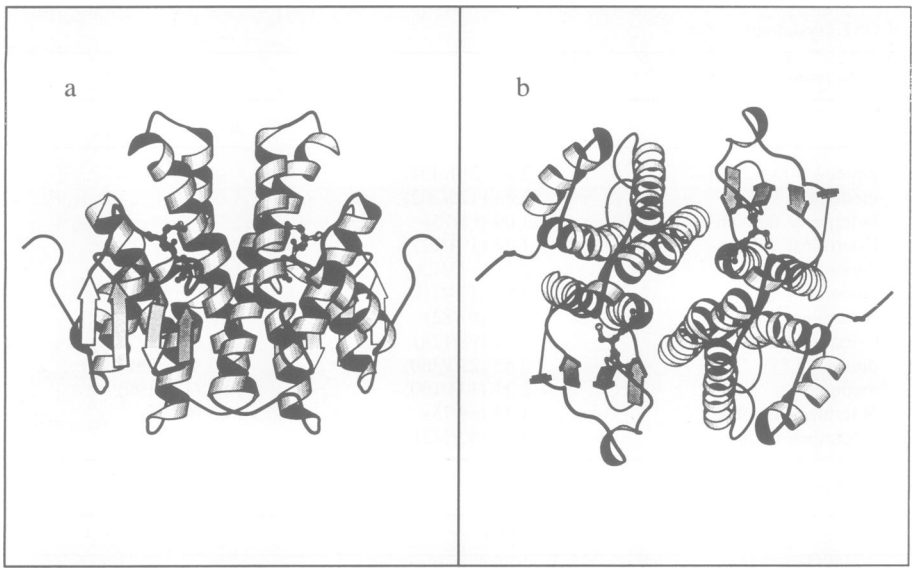

Fig. 4. Ribbon representation of *L.cuprina* GST dimer. Residues were assigned to secondary structural elements according to criteria defined by Kabsch and Sander (1983). The bound GSH is shown in ball-and-stick fashion. Two orthogonal views are shown. The figures were produced by the computer program MOLSCRIPT (Kraulis, 1991).

bridge (Glu 74 to Arg 90). All these interactions are close to the 2-fold axis and form two layers of interaction: one layer at the base of the V-shaped cleft consisting of the salt bridge and the Tyr 71–Lys 88 interaction, and the remainder of the interactions occupying the second layer located half-way up the V-shaped cleft. The majority of interactions between the monomers consist of van der Waals interactions and an extensive array of water-mediated contacts. Of particular note is the parallel stacking of symmetry-related aromatic rings of Tyr 98 located in the second interaction layer. The conserved intersubunit contact created by the stacking of two symmetrically equivalent arginine guanidino groups in the mammalian enzymes (Sinning *et al.*, 1993; Dirr *et al.*, 1994) is not observed in the *Lucilia* structure.

The active site

The active site is located in a deep (20 Å) V-shaped cleft formed by the abutting of the two domains (Figures 1 and 4). The cleft is ~5 Å deeper than that observed in the mammalian GST crystal structures due to the extended loop structure between helices 4 and 5 in *Lucilia* GST (Figure 5). The active site of each monomer is well separated in the dimer with the closest atomic distance between inhibitor molecules of 13.5 Å (Figure 4b). We do not observe density for the hexyl portion of the inhibitor (Figure 6). Because the enzyme is eluted with GSH from a GSH–agarose column during the purification procedure, it is possible that we have crystallized a GSH complex rather than a hexyl-GSH complex. In the human mu-class structure, electron density is observed only for the GSH portion of the dinitrobenzene–GSH inhibitor (Raghunathan *et al.*, 1994). The authors suggest that the absence of density may be due to either high mobility or enzymatic cleavage of the aromatic ring. Details of the GSH interactions with the protein are shown in Figures 7 and 8. There are a total of 15 potential salt links and hydrogen bonds with the GSH moiety. Two water molecules are observed within hydrogen-bonding distance of the thiol sulfur. Similarly located water molecules are also seen in some of the other GST crystal structures (García-Sáez *et al.*, 1994; Ji *et al.*, 1992; Raghunathan *et al.*, 1994). Superposition of the various crystal structures demonstrates that the GSH molecule is orientated in almost identical fashion with the γ-glutamyl arm pointing downward toward the dimer interface and the glycyl portion pointing away from the core of the N-terminal domain towards solvent (Figure 9). Unlike the mammalian GSTs, there are no residues contributed from the neighbouring subunit to the active site. In contrast to the mammalian crystal structures, the *Lucilia* crystal structure is remarkable for its very open hydrophobic binding site which can be described as a deep V-shaped pocket that is fully exposed to solvent (Figures 3 and 5). The pocket is lined with residues Leu 6–Ser 9, Leu 33, Tyr 105, Tyr 113 and Phe 117. The base of the pocket is lined with polar atoms, including the main-chain atoms of Pro 7, Gly 8 and the hydroxyl group of Ser 9. The polar base and high solvent exposure of the pocket suggest that the *Lucilia* enzyme is not as well adapted to the binding of hydrophobic substrate compared with its mammalian counterparts.

Catalytic mechanism

Calculations suggest that the thiolate anion is up to 10^9 times more reactive than its conjugate acid (Roberts *et al.*, 1986). There are two possibilities of how the activation may occur in GSTs. One possibility is polarization of the thiol S–H bond by hydrogen bonding to a suitable base. Another possibility is ionization to form the thiolate anion prior to conjugation. There is considerable evidence in support of the latter possibility for the mammalian enzymes (Armstrong, 1991). Numerous studies of mammalian GSTs have demonstrated that residues involved in GSH binding and activation are almost solely contributed from the N-terminal domain (Reinemer *et al.*, 1991, 1992; Ji *et al.*, 1992; Sinning *et al.*, 1993; García-Sáez *et al.*, 1994; Raghunathan *et al.*, 1994). The importance of the N-terminal domain is reflected by the higher sequence

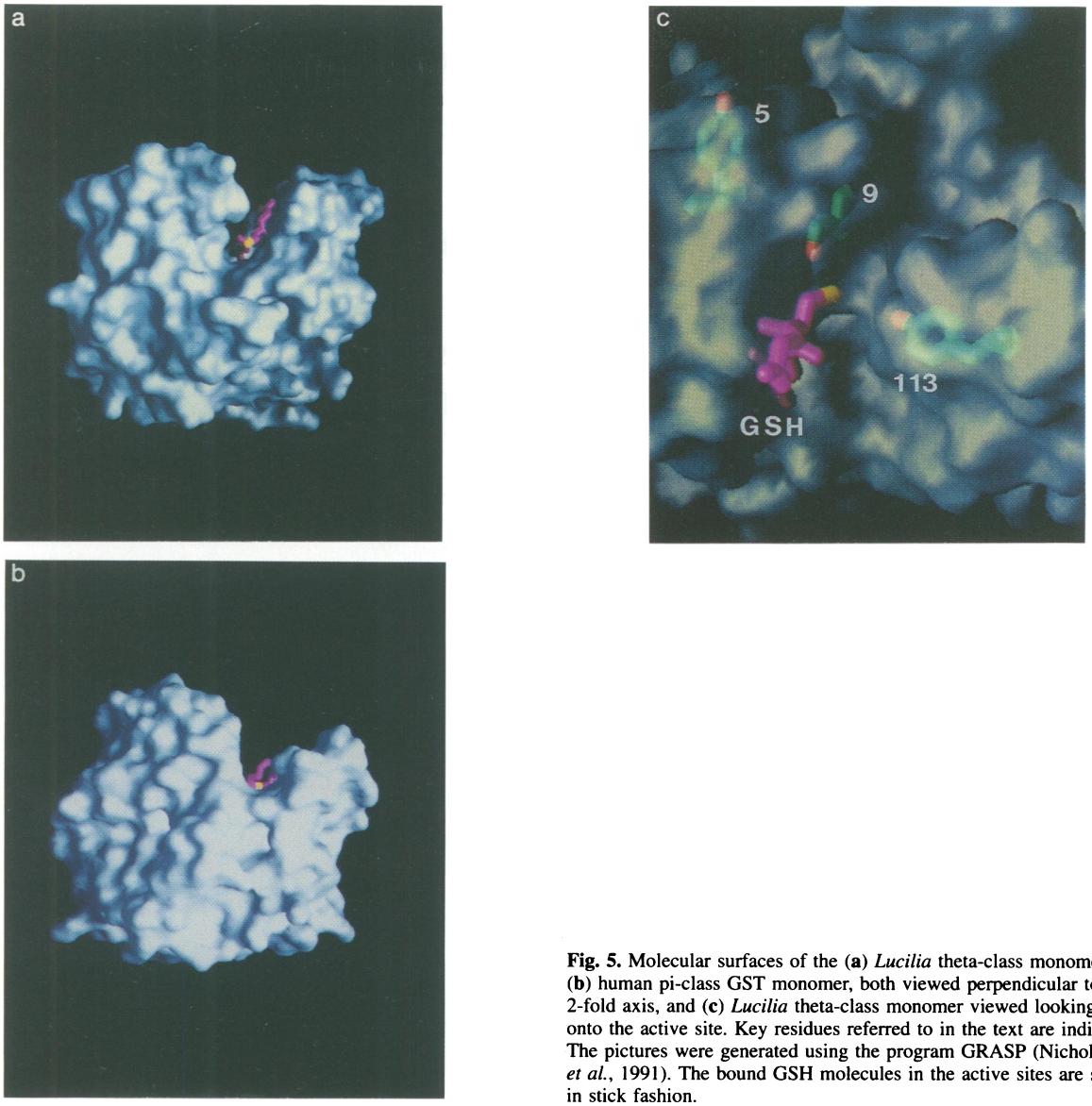

Fig. 5. Molecular surfaces of the (**a**) *Lucilia* theta-class monomer, (**b**) human pi-class GST monomer, both viewed perpendicular to the 2-fold axis, and (**c**) *Lucilia* theta-class monomer viewed looking down onto the active site. Key residues referred to in the text are indicated. The pictures were generated using the program GRASP (Nicholls *et al.*, 1991). The bound GSH molecules in the active sites are shown in stick fashion.

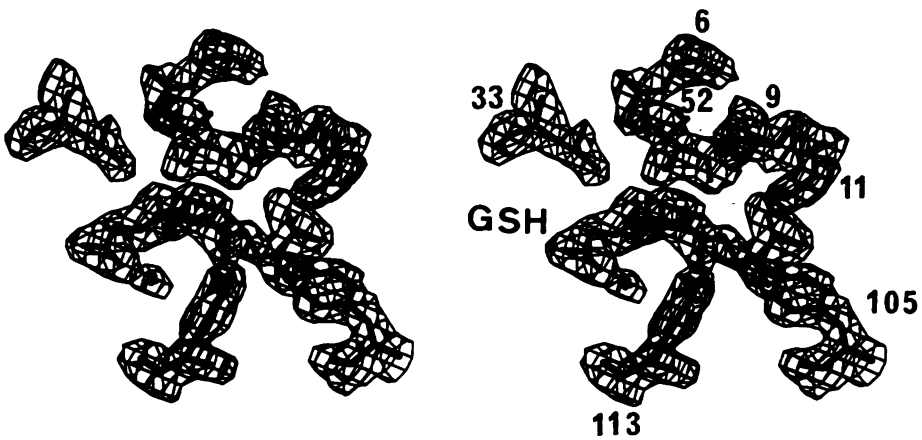

Fig. 6. Stereo view of a 2.2 Å resolution $2F_o - F_c$ Fourier omit map of the active site region (where the residues and water molecules in the active site and GSH were omitted from the Fourier calculation).

identity of this domain among GSTs and lower r.m.s. deviations of alpha carbon positions of this domain on superposition of the known crystal structures (Tables I and II) (Wilce and Parker, 1994). Of particular importance is the observation that a tyrosine near the N-terminus is conserved in almost all published sequences of GSTs (see Figure 2). The importance of this residue for activation of GSH has been demonstrated in numerous mutagenesis

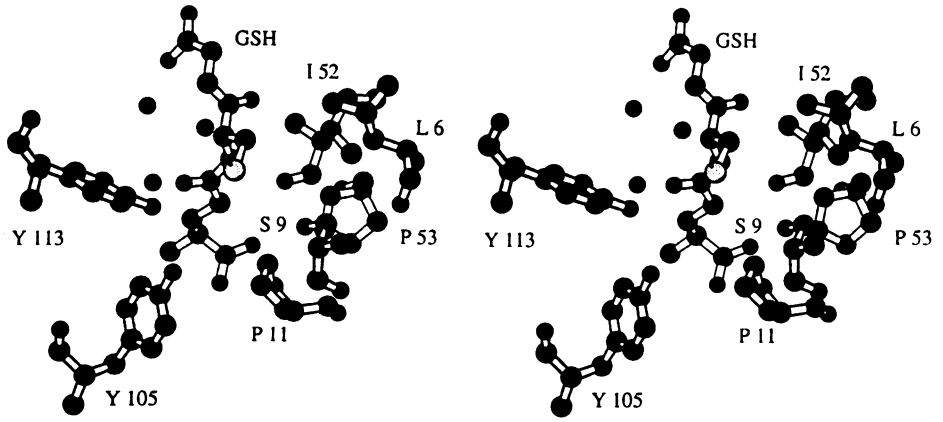

Fig. 7. Stereo view of the active site generated using the program MOLSCRIPT (Kraulis, 1991).

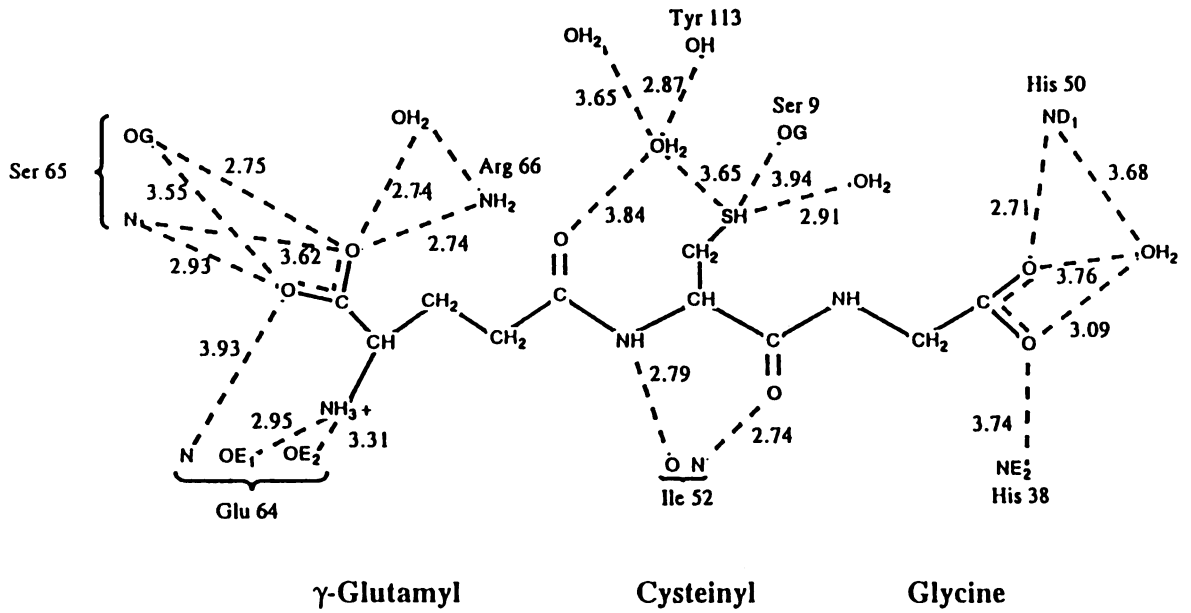

Fig. 8. Schematic drawing of residues in *Lucilia* GST that interact with GSH. Potential hydrogen bonds are indicated by broken lines and distances between polar atoms shown are in units of Ångstroms.

studies of the alpha-, mu- and pi-class enzymes (Stenberg *et al.*, 1991; Kolm *et al.*, 1992; Kong *et al.*, 1992a,b; Liu *et al.*, 1992; Manoharan *et al.*, 1992a,b; Penington and Rule, 1992; Wang *et al.*, 1992). The activation of the thiol group is achieved by lowering the pK_a of the thiol group to generate a thiolate anion (Armstrong, 1991). Recent evidence suggests that this tyrosine can dissociate to some extent to form a tyrosinate ion at physiological pH, and thus may act as a general base for abstracting the proton from the GSH thiol (Atkins *et al.*, 1993; Meyer *et al.*, 1993). It is thus surprising to find that this tyrosine, although present in the *Lucilia* enzyme, is nowhere near the active site in the crystal structure. The hydroxyl of Tyr 5 is 13.9 Å distant from the thiol sulfur atom (Figure 5c). [The sequence interpretation in this region of the electron density map is unambigous (Figure 10). The density for the tyrosine residue is excellent with a real space correlation coefficient (Jones *et al.*, 1991) of 0.94. The average temperature factors for the the main-chain and side-chain atoms of this residue are 15.1 and 20.3 $Å^2$, respectively.]

Assuming that the pK_a of the GSH thiol group is lowered in the *Lucilia* enzyme, there are two alternative residues that may be responsible for the activation. On superposition of the GST crystal structures, the hydroxyl of Ser 9 superimposes close to the position of the catalytically important tyrosine of the mammalian enzymes. However, the hydroxyl of Ser 9 is 3.9 Å away from the GSH sulfur atom. Another candidate residue is Tyr 113, which could hydrogen bond to the GSH thiol sulfur through a water molecule (Figures 5c–8). There is an intricate network of potential hydrogen bonding in this region with two water molecules within hydrogen-bonding distance of the thiol sulfur and each other. Furthermore, Tyr 105 is within hydrogen-bonding distance of Tyr 113. Both Ser 9 and Tyr 113 are highly exposed to solvent, unlike the catalytic

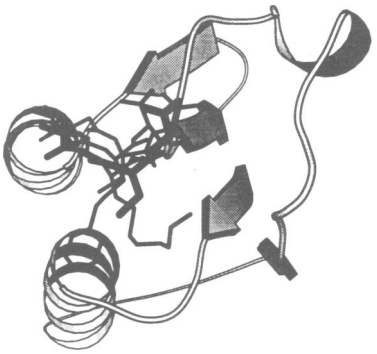

Fig. 9. A comparison of the binding locations of GSH conjugates derived from the various crystal structures. The superposition was based on overlaying the coordinates of the respective N-terminal domains. The ligands are overlayed on a schematic diagram of the N-terminal domain of the *L.cuprina* enzyme. The GSH conjugates are shown in stick fashion with the one from the *Lucilia* enzyme highlighted with thickened lines. The figure was produced by the computer program MOLSCRIPT (Kraulis, 1991).

tyrosine in the mammalian enzymes which is mostly buried. It is likely that the burying of the latter residue prevents potential hydrogen-bond acceptors from interacting with the tyrosine hydroxyl and thus enhances its hydrogen-bond donor capacity towards the sulfur atom of bound GSH. Only recently has evidence come to light that residues from the C-terminal domain may be directly involved in catalysis. The tyrosine residue equivalent to Tyr 113 in a mu-class GST has been implicated in GSH-dependent epoxide ring opening and Michael additions (Barycki and Colman, 1993; Johnson *et al.*, 1993; Ji *et al.*, 1994). The hydroxyl of this tyrosine is positioned so as to act as a general acid catalytic group. The same tyrosine appears to play a role in the stabilization of a σ-complex formed in the reaction of the mu-class GST with 1-chloro-2,4-dinitrobenzene (Ji *et al.*, 1993). However, although conserved in all known pi- and mu-class sequences, this tyrosine is not conserved in the human alpha-class enzyme, where it is replaced by a hydrophobic residue. Its potential role in activating GSH appears unique to the theta-class enzymes. It should be pointed out that Tyr 113, unlike Ser 9, is not an invariant residue in the theta-class family (Figure 2). The relative importance of each residue in the activation of GSH and catalysis is currently under investigation by site-directed mutagenesis.

Concluding remarks

Most theta-class GSTs are difficult to purify because they are not quantitatively retained on GSH-based affinity columns (Meyer *et al.*, 1991). *Lucilia* GST is only weakly bound (Board *et al.*, 1994). An explanation for this difficulty was provided by a recent observation that a long spacer arm of 21 carbons linked through the thiol group of GSH was successfully used in purifying a plant theta-class GST (Lopez *et al.*, 1994). In contrast, many commercially available resins use shorter spacers (typically 10 carbons) which are often attached to the GSH amino group. These observations are readily understood in terms of the crystal structure presented here. The active site of the theta-class enzyme is much deeper than that found in the mammalian crystal structures and the amino group is buried at the bottom of the active site close to the monomer–monomer interface. Conjugation of the spacer arm through the sulfur atom has clear advantages, given that the thiol moiety is highly exposed to solvent. The likely extension of helix 5 in the mammalian theta-class GSTs suggests that the cleft could be even deeper in these enzymes.

The crystal structures and amino acid sequence alignments have highlighted the importance of key theta-class residues. For example, the following invariant residues appear of critical importance in determining the GST fold: Pro 53, Gly 149 and Asp 156. The serine near the N-terminus (Ser 9 in *Lucilia*) and an asparagine close to the invariant proline (Asn 47 in *Lucilia*) appear to distinguish theta-class GSTs from the other GST families (Figure 2 and M.C.J.Wilce and M.W.Parker, unpublished results of more extensive sequence alignments). The serine has been implicated in the catalytic mechanism (see above). The asparagine residue is buried and is involved in extensive hydrogen-bonding interactions between helix 2 and the rest of the N-terminal domain, suggesting it might play an important role in the positioning of this helix relative to the active site (see Figure 1). The asparagine side chain makes contacts with the hydroxyl groups of Thr 51 and Thr 54, and main-chain atoms of Gln 49, His 50 and Thr 51. However, none of the contacting residues are totally conserved in the theta-class sequences. It is clear from the crystal structure of *Lucilia* GST that the presence of a tyrosine near the N-terminus is no longer an obligatory condition in order to classify a protein as a GST.

Materials and methods

Crystals of *L.cuprina* GST were grown in the presence of the inhibitor, *S*-hexyl GSH, as described elsewhere (Wilce *et al.*, 1994). The crystals belong to the tetragonal space group $P4_322$ with cell dimensions of $a = b = 88.9$ Å and $c = 65.9$ Å. Although the biggest crystals were only 0.15×0.15×0.15 mm, diffraction can be measured to beyond 2.2 Å resolution. Data used in the structure determination were collected using a MARRESEARCH area detector with CuKα X-rays generated by a Rigaku RU-200 rotating anode generator. All data were collected at 18°C. The diffraction data were processed and analysed using programs in the XDS (Kabsch, 1988) and CCP4 (1994) suites. The relatively high R_{merge} values in Table III reflect the small size of the available crystals. The R_{merge} value for the highest resolution shell (2.3–2.2 Å) was 37.2% and 73% of the reflections in this resolution bin had intensity measurements >3 SD. Attempts to solve the structure by molecular replacement methods using the available mammalian GST crystal structures proved unsuccessful. A single, significant signal to the rotation function was found irrespective of which GST model was used and which computer program was employed. The best solution was obtained using the alpha-class monomer as a search model and the Crowther fast rotation function [program ALMN of the CCP4 package (1994)]. A peak of 5.9σ was obtained using a 25 Å sphere in the resolution range 6–4 Å. The rotation function result was later shown to be correct once the structure was solved. However, finding a solution to the translation problem proved intractable, despite using a variety of Patterson and *R*-factor searching techniques available in the XPLOR (Brünger *et al.*, 1987) and CCP4 (1994) packages. We suspect that the failure of the translation function to work is due to the significant differences,

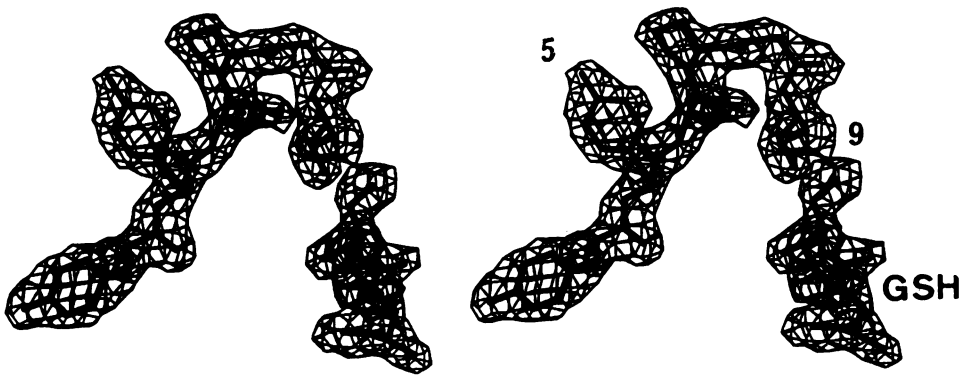

Fig. 10. Stereo view of a 2.2 Å resolution $2F_o - F_c$ Fourier omit map of the region about Tyr 5 (where the tyrosine and four residues on either side of it were omitted from the Fourier calculation). The figure was produced with the program O (Jones *et al.*, 1991).

Table III. Summary of diffraction data and phasing statistics

Data set	Native	$KAu(CN)_2$	$Hg(CH_3COO)_2$
Soak time (h)	–	16	16
Soaking concentrations (mM)	–	5	1
Resolution (Å)	2.2	3.5	3.5
No. of observations	49 556	7128	8892
No. of unique reflections	13 137	3093	3175
Data completeness (%)	92	88	96
No. of data $>3\sigma_I$ (%)	76	84	81
R_{merge}[a] (%)	12.6	13.4	14.7
MFID[b] (%)	–	18.2	18.3
No. of sites	–	2	2
R_{Cullis}[c] (%)	–	74.5	69.5
$<\|F_H\|>/E$[d]	–	1.4	1.3

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i - <I>|/|<I>|$, where I_i is the intensity for the ith measurement of an equivalent reflection with indices h, k, l.
[b] $MFID = \Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$ where F_{PH} refers to derivative data and F_P to native data.
[c] $R_{Cullis} = \Sigma||F_{PHC}| - |F_{PH}||/\Sigma||F_{PH}| - |F_P||$. The summation is over all centric reflections. F_{PHC} and F_{PH} are calculated and measured derivative structure factor amplitudes, respectively. F_P is the native structure factor.
[d] $<|F_H|>/E$ is a measure of the phasing power of the derivative where $<|F_H|>$ is the r.m.s. heavy atom derivative structure factor amplitude and E is the lack-of-closure error.

particularly in the helix lengths and helix curvatures, between the *Lucilia* GST model and the other GST crystal structures. Heavy atom derivatives were obtained by soaking crystals in artificial mother liquor containing the derivative at room temperature (see Table III). Two sites were located in the isomorphous difference Patterson function of the mercury derivative and subsequent sites were found by cross-difference Fourier analysis. Anomalous scattering data were collected for both derivatives and were used to establish the correct enantiomorphic space group unequivocally. Heavy atom parameters were refined using MLPHARE (CCP4 Suite, 1994). The overall figure of merit was 0.42 (for the resolution shell 20.0–3.8 Å). The initial MIR map was clear enough to identify elements of the canonical GST polypeptide fold, but required improvement to enable the amino acid sequence to be built into the map. The map was improved with the program SQUASH using MIR phases to 3.8 Å resolution (Cowtan and Main, 1993). Use was made of solvent flattening, histogram matching and Sayre's equation to extend the phases to 2.2 Å resolution. The starting model was built into the SQUASH map on an Evans and Sutherland ESV3 graphics workstation using the program O (Jones *et al.*, 1991). Partial models were improved with molecular dynamics refinement using the slowcool protocol of XPLOR (Brünger *et al.*, 1987). Phases from these models were combined with MIR and SQUASH phases using COMBINE (CCP4 Suite, 1994). In all, nine rounds of model building and MD refinement were required to produce the final model. The map interpretation was checked using residue omit maps. About 10% of the model was deleted, a few cycles of MD refinement performed to reduce bias, and the resultant $2F_o - F_c$ electron density map examined in the region of omission. The correctness of the interpretation was also confirmed by the location of two mercury binding sites at Cys 12 and Cys 86. The final crystallographic R-factor for all data in the resolution range 6.0–2.2 Å (12 190 reflections) is 17.9% (R_{free} = 24.3%) for the model consisting of residues 1–201 and 129 water molecules. The N-terminal methionine found in many GSTs (Figure 2) is present in the expressed *Lucilia* enzyme, as judged by N-terminal peptide sequencing (data not shown). The r.m.s. deviations of the model from ideal geometry are 0.007 Å for bond lengths and 1.9° for bond angles. The r.m.s. deviation for dihedrals is 23.9° and the r.m.s. on impropers is 1.22°. The r.m.s. deviations on bonded atoms are 21.7 $Å^2$ for main-chain atoms and 24.2 $Å^2$ for side-chain atoms. There are no outliers in the Ramachandran plot, with the exception of Cys 201 which lies in poorly resolved electron density in the final map. The path of the polypeptide chain is clear and most of the side chain density is well defined, with the exception of a few charged surface residues. The final electron density map suggested that Thr 10, Tyr 12 and His 13 had been sequenced incorrectly, possibly due to compressions in the nucleotide sequence. This has since been confirmed through further DNA sequencing (P.G.Board, unpublished results). N-Terminal peptide sequencing analysis demonstrated that these residue positions are occupied by Ala, Cys and Arg, respectively, and thus the published cDNA sequence (Board *et al.*, 1994) requires correction accordingly. The coordinates will be deposited with the Brookhaven Protein Databank (Bernstein *et al.*, 1977).

Acknowledgements

We thank Ken Mitchelhill for carrying out the N-terminal peptide sequencing. M.W.P. is a Wellcome Australian Senior Research Fellow, and acknowledges support from the Australian Research Council and the National Health and Medical Research Council. Financial support from the Australian National Beamline Facility is gratefully acknowledged.

References

Armstrong,R.N. (1991) *Chem. Res. Toxicol.*, **4**, 131–140.

Atkins,W.M., Wang,R.W., Bird,A.W., Newton,D.J. and Lu,A.Y.H. (1993) *J. Biol. Chem.*, **268**, 19188–19191.

Bairoch,A. and Boeckmann,B. (1991) *Nucleic Acids Res.*, **19**, 2247–2249.

Barycki,J.J. and Colman,R.F. (1993) *Biochemistry*, **32**, 13002–13011.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Board,P., Coggan,M., Johnston,P., Ross,V., Suzuki,T. and Webb,G. (1990) *Pharm. Ther.*, **48**, 357–369.

Board,P.G., Russell,R.J., Marano,R.J. and Oakeshot,J.G. (1994) *Biochem. J.*, **299**, 425–430.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) *Science*, **235**, 458–460.

Buetler,T.M. and Eaton,D.L. (1992) *Environ. Carcinogen. Ecotoxicol. Rev.*, **C10**, 181–203.

CCP4 Suite (1994) *Acta Crystallogr. Sect. D*, **50**, 750–763.

Clark,A.G. (1989) *Comp. Biochem. Physiol.*, **92b**, 419–446.

Cowtan,K.D. and Main,P. (1993) *Acta Crystallogr. Sect. D*, **49**, 148–157.

Dirr,H., Reinemer,P. and Huber,R. (1994) *Eur. J. Biochem.*, **220**, 645–661.

Fournier,D., Bride,J.M., Poirie,M., Bergé,J.-B. and Plapp,F.W. (1992) *J. Biol. Chem.*, **267**, 1840–1845.

García-Sáez,I., Párraga,A., Phillips,M.F., Mantle,T.J. and Coll,M. (1994) *J. Mol. Biol.*, **237**, 298–314.

Hiratsuka,A. *et al.* (1990) *J. Biol. Chem.*, **265**, 11973–11981.

Ji,X., Zhang,P., Armstrong,R.N. and Gilliland,G.L. (1992) *Biochemistry*, **31**, 10169–10184.

Ji,X., Armstrong,R.N. and Gilliland,G.L. (1993) *Biochemistry*, **32**, 12949–12952.

Ji,X., Johnson,W.W., Sesay,M.A., Dickert,L., Prasad,S.M., Ammon,H.L., Armstrong,R.N. and Gilliland,G.L. (1994) *Biochemistry*, **33**, 1043–1052.

Johnson,W.W., Liu,S., Ji,X., Gilliland,G.L. and Armstrong,R.N. (1993) *J. Biol. Chem.*, **268**, 11508–11511.

Jones,T.A., Zhou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr. Sect. A*, **47**, 110–119.

Kabsch,W. (1988) *J. Appl. Crystallogr.*, **21**, 916–924.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Kolm,R.H., Sroga,G.E. and Mannervik,B. (1992) *Biochem. J.*, **285**, 537–540.

Kong,K.-H., Nishida,M., Inoue,H. and Takahashi,K. (1992a) *Biochem. Biophys. Res. Commun.*, **182**, 1122–1129.

Kong,K.-H., Takasu,K., Inoue,H. and Takahashi,K. (1992b) *Biochem. Biophys. Res. Commun.*, **184**, 194–197.

Kraulis,J.P. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Liu,S., Zhang,P., Ji,X., Johnson,W.W., Gilliland,G.L. and Armstrong,R.N. (1992) *J. Biol. Chem.*, **267**, 4296–4299.

Lopez,M.F., Patton,W.F., Sawlivich,W.B., Erdjument-Bromage,H., Barry,P., Gmyrek,K., Hines,T., Teempst,P. and Skea,W.M. (1994) *Biochim. Biophys. Acta*, **1205**, 29–38.

Mannervik,B. and Danielson,U.H. (1988) *CRC Crit. Rev. Biochem.*, **23**, 283–337.

Mannervik,B. *et al.* (1993) In Pickett,K.D., Mantle,T.J., Mannervik,B. and Hayes,J.D. (eds), *Structure and Function of Glutathione Transferases*. CRC Press, Boca Raton, FL, pp. 29–38.

Manoharan,T.H., Gulick,A.M., Puchalski,R.B., Servais,A.L. and Fahl,W.E. (1992a) *J. Biol. Chem.*, **267**, 18940–18945.

Manoharan,T.H., Gulick,A.M., Reinemer,P., Dirr,H.W., Huber,R. and Fahl,W.E. (1992b) *J. Mol. Biol.*, **226**, 319–322.

Meyer,D.J., Coles,B., Pemble,S.E., Gilmore,K.S., Fraser,G.M. and Ketterer,B. (1991) *Biochem. J.*, **274**, 409–414.

Meyer,D.J., Xia,C., Coles,B., Chen,H., Reinemer,P., Huber,R. and Ketterer,B. (1993) *Biochem. J.*, **293**, 351–356.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) *Protein Struct. Funct. Genet.*, **11**, 281–293.

Pemble,S.E. and Taylor,J.B. (1992) *Biochem. J.*, **287**, 957–963.

Penington,C.J. and Rule,G.S. (1992) *Biochemistry*, **31**, 2912–2920.

Raghunathan,S., Chandross,R.J., Kretsinger,R.H., Allison,T.J., Penington,C.J. and Rule,G.S. (1994) *J. Mol. Biol.*, **238**, 815–832.

Reinemer,P., Dirr,H.W., Ladenstein,R., Schäffer,J., Gallay,O. and Huber,R. (1991) *EMBO J.*, **10**, 1997–2005.

Reinemer,P., Dirr,H.W., Ladenstein,R., Huber,R., Lo Bello,M., Federici,G. and Parker,M.W. (1992) *J. Mol. Biol.*, **227**, 214–226.

Roberts,D.D., Lewis,S.D., Ballou,D.P., Olson,S.T. and Shafer,J.A. (1986) *Biochemistry*, **25**, 5595–5601.

Sinning,I. *et al.* (1993) *J. Mol. Biol.*, **232**, 192–212.

Stenberg,G., Board,P.G. and Mannervik,B. (1991) *FEBS Lett.*, **293**, 153–155.

Wang,R.N., Newton,D.J., Huskey,S.W., McKeever,B.M., Pickett,C.B. and Lu,A.Y.H. (1992) *J. Biol. Chem.*, **267**, 19866–19871.

Wilce,M.C.J. and Parker,M.W. (1994) *Biochim. Biophys. Acta*, **1205**, 1–18.

Wilce,M.C.J., Feil,S.C., Board,P.G. and Parker,M.W. (1994) *J. Mol. Biol.*, **236**, 1407–1409.

Received on November 22, 1994; revised on February 16, 1995