Molecular basis of glutathione synthetase deficiency and a rare gene permutation event

**Galina Polekhina, Philip G.Board[1],
Rayappa R.Gali[1], Jamie Rossjohn 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]Molecular Genetics Group, Division of Molecular
Medicine, John Curtin School of Medical Research, Australian
National University, GPO Box 334, Canberra 2601, Australia

[2]Corresponding author
e-mail: mwp@rubens.its.unimelb.edu.au

Glutathione synthetase (GS) catalyses the production
of glutathione from γ-glutamylcysteine and glycine in
an ATP-dependent manner. Malfunctioning of
GS results in disorders including metabolic acidosis,
5-oxoprolinuria, neurological dysfunction, haemolytic
anaemia and in some cases is probably lethal. Here we
report the crystal structure of human GS (hGS) at
2.1 Å resolution in complex with ADP, two magnesium
ions, a sulfate ion and glutathione. The structure
indicates that hGS belongs to the recently identified
ATP-grasp superfamily, although it displays no detect-
able sequence identity with other family members
including its bacterial counterpart, *Escherichia coli*
GS. The difficulty in identifying hGS as a member of
the family is due in part to a rare gene permutation
which has resulted in a circular shift of the conserved
secondary structure elements in hGS with respect to
the other known ATP-grasp proteins. Nevertheless, it
appears likely that the enzyme shares the same general
catalytic mechanism as other ligases. The possibility
of cyclic permutations provides an insight into the
evolution of this family and will probably lead to the
identification of new members. Mutations that lead to
GS deficiency have been mapped onto the structure,
providing a molecular basis for understanding their
effects.
Keywords: ATP-grasp superfamily/genetic disease/
glutathione/glutathione synthetase/X-ray crystallography

Introduction

Glutathione (γ-glutamylcysteinylglycine, GSH) is the most
abundant intracellular thiol in living aerobic cells. It has
been assigned several critical functions: protection of cells
against oxidative damage; involvement in amino acid
transport; participation in the detoxification of foreign
compounds; maintenance of protein sulfhydryl groups in
a reduced state; and as a cofactor for a number of enzymes.
Low GSH levels have been associated with the pathology
of a number of diseases including AIDS, hepatitis C,
type II diabetes, ulcerative colitis, idiopathic pulmonary
fibrosis, adult respiratory distress syndrome, cataracts

and neurological disorders (Anderson, 1998). GSH is
synthesized from glutamate, cysteine and glycine by the
consecutive action of two enzymes: γ-glutamylcysteine
synthetase (γGCS, EC 6.3.2.2) and GSH synthetase (GS,
EC 6.3.2.3). Together with γ-glutamyl transpeptidase,
γ-glutamyl cyclotransferase and 5-oxoprolinase, these
enzymes constitute the γ-glutamyl cycle. GSH acts as a
feedback inhibitor of the cycle because it can competitively
inhibit γGCS (Ristoff and Larsson, 1998).

Genetic studies have revealed mutations in the human
GS gene which lead to GS deficiency and suggest that
complete loss of function is probably lethal (Shi *et al.*,
1996; Dahl *et al.*, 1997). Although there appears to be
only one GS gene (Webb *et al.*, 1995), there are two
forms of the disease (Ristoff and Larsson, 1998): a milder
form, caused by a GS deficiency limited to erythrocytes,
resulting in chronic haemolytic anaemia and neonatal
jaundice; and a more severe form which additionally leads
to 5-oxoprolinuria and metabolic acidosis. Several patients
suffering from the severe form of the disease are mentally
retarded and some exhibit motor function disturbances
(Ristoff and Larsson, 1998). In general, GS deficiency
leads to an overproduction of γ-glutamylcysteine due to a
breakdown of feedback inhibition by the lowered levels
of GSH. The excess γ-glutamylcysteine is converted by
γ-glutamyl cyclotransferase into 5-oxoproline and cysteine.
The excessive production of 5-oxoproline exceeds the
capacity of 5-oxoprolinase, and 5-oxoproline accumulates
in the body fluids, leading to metabolic acidosis and
5-oxoprolinuria (Meister and Larsson, 1995).

Human GS (hGS) has received much attention because
of its involvement in hereditary disease. The human cDNA
and gene have been cloned and sequenced (Gali and
Board, 1995; Whitbread *et al.*, 1998) and the primary
structures of GS from a number of eukaryotes are also
known. These studies indicate that hGS is a homodimer
of 52 kDa subunit molecular weight and it shares amino
acid sequence identity of between 35 and 89% with
the other eukaryotic enzymes (Whitbread *et al.*, 1998),
whereas there is no detectable sequence similarity between
the eukaryotic enzymes and their bacterial counterpart.
The *Escherichia coli* enzyme (ecGS) is 158 residues
shorter than hGS and exists as a tetramer rather than a
dimer (Gushima *et al.*, 1984). The crystal structure of
ecGS has been determined (Yamaguchi *et al.*, 1993) and
represents the first member of what is now referred to as
the ATP-grasp superfamily (Murzin, 1996). A common,
defining feature of the family is that they all exhibit
carboxylate-amine/thiol ligase activity despite acting on a
widely diverse range of substrates. Their structures are
characterized by an ATP-binding cleft formed by two anti-
parallel sheets and a phosphate-binding loop (Galperin
and Koonin, 1997). Other members include D-alanine:D-
alanine ligase (DDL; Fan *et al.*, 1994), biotin carboxylase

α-chain (BC; Waldrop *et al.*, 1994), succinyl CoA synthetase β-chain (SCS; Wolodko *et al.*, 1994), carbamoyl phosphate synthase (CPS; Thoden *et al.*, 1997) and the C-domain of synapsin I (Esser *et al.*, 1998). The superfamily can be divided into two subfamilies: those that possess three common domains (ecGS, DDL, BC, CPS and synapsin) and those where the similarity is restricted to two domains (SCS). Despite all these proteins sharing the same basic fold, there is barely any detectable amino acid sequence identity (~10–20%) between them. Nevertheless, Galperin and Koonin (1997) were able to expand the list of family members to 15 by searching sequence databases with motifs centred around a flexible glycine-rich loop that forms part of the ATP-binding site. Until this study it was not known whether hGS was a member of the ATP-grasp superfamily: based on function it should be a member but lack of any detectable sequence identity to known members, particularly to its bacterial counterpart, has raised doubts.

We have determined the crystal structure of hGS in complex with ADP, GSH, magnesium and sulfate ions by multiple isomorphous replacement and refined the model to 2.1 Å resolution. The structure represents the enzyme–product complex with one of the sulfate ions mimicking the inorganic phosphate after the ligation reaction has occurred. The structure reveals that hGS does belong to the ATP-grasp superfamily but surprisingly the conserved secondary structure elements have been circularly permuted. We argue that this is the result of a rare gene permutation event that occurred before eukaryotes evolved.

Results

Overall structure

The final model of hGS comprises residues 3–474, ADP, two magnesium ions, two sulfate ions and 230 water molecules. hGS is a compact molecule with the shape of a flat, equilateral triangle with the sides of the triangle of ~60 Å and thickness of ~40 Å (Figure 1A and B). The ligands, consisting of ADP, magnesium ions, GSH and one of the sulfate ions, are bound in a central cavity on one side of the molecule, with ADP stacked between two anti-parallel β-sheets. The cavity is covered by three loops (residues 266–276, 366–374, 454–466; designated gray in Figure 1B) projecting from three of the main structural units of the structure. The first loop forms interactions with the GSH so we refer to it as the substrate-binding loop. The other two loops have been named the Gly-rich loop and Ala-rich loop, respectively, because the nomenclature reflects their amino acid composition. The main structural units are an anti-parallel β-sheet (residues 48–184, 402–474; strands β3, β4, β5, β16, β17, β18 and β19; designated blue in Figure 1B), together with helices (α3, α4, α5, α6, α7) packing on either side of the sheet, a parallel β-sheet (residues 185–294, strands β6, β7, β8, β9, β10, β11; designated red in Figure 1B) together with helices (α8, α9, α10, α12) on both sides and a domain we call the lid because of its role in providing access to the ATP-binding site (see below). The lid domain (residues 336–401; designated green in Figure 1B) consists of an anti-parallel sheet (strands β12, β13, β14, β15) with helices (α17, α18, α19) packed on one side. In addition

to the three main structural units, residues 3–48 form a dimerization unit (helices α1, α2 and strands β1, β2; designated purple in Figure 1B) and residues 295–335 make up the helical connection (helices α13, α14, α15 and α16; designated orange in Figure 1B) between the parallel β-sheet unit and the lid domain. A mini-β barrel is located between the dimerization unit and the edge of the anti-parallel β-sheet unit. The barrel is built from portions of five strands, β3, β16, β17, β18 and β19, and a sulfate ion is bound at one end (Figure 1B).

ATP-binding site

ADP binds to hGS in a manner similar to that observed in other ATP-grasp proteins (Fan *et al.*, 1994; Waldrop *et al.*, 1994; Wolodko *et al.*, 1994; Thoden *et al.*, 1997; Esser *et al.*, 1998). It is sandwiched between the strands of two anti-parallel β-sheets: strand pairs β4, β5 and β13, β14 (Figure 1B). The adenine-binding pocket is largely hydrophobic with contributions from Met129, Ile143, Val362, Met398, Ile401 and the aliphatic portions of Lys364 and Lys400. The main-chain carbonyl oxygen of Glu399 can form a hydrogen bond to the N6 atom of the adenine base and the amide of Ile401 can interact with the N1 nitrogen. The hydroxyl of Tyr375, $O_\varepsilon2$ of Glu425 and the NZ of Lys452 are within hydrogen bonding distance of the O4′, O3′ and O2′ atoms of the ribose, respectively (Figure 2). The negative charges on the α- and β-phosphates are compensated by the positively charged residues, Lys305 and Lys364 (Figure 2). In addition, there are polar interactions between the phosphate oxygens and Asn373 and the main-chain amide of Gly370.

Magnesium binding

Two magnesium ions (Mg1 and Mg2) have been located in the structure (Figure 2) and both are bound in an octahedral geometry. The ligands around Mg1 are an oxygen atom from the α-phosphate of ADP, an oxygen atom from the β-phosphate of ADP, an oxygen atom from a sulfate ion, a carboxylate oxygen of Glu144 and two water molecules. The metal–ligand distances range between 2.1 and 2.2 Å. Mg2 is ligated by an oxygen atom from the β-phosphate of ADP, a carboxylate oxygen from Glu368, Oδ1 of Asn146, both carboxylate oxygens of Glu144 and an oxygen atom of a sulfate ion. The bond distances between metal and ligand vary between 1.9 and 2.4 Å. Glu144 seems a particularly important residue as it forms a bridging interaction between the two metal sites.

The sulfate ion binding in the active site

Although the protein was crystallized in the presence of ATP, no density for the γ-phosphate was observed. It is very likely that the ATP hydrolysed spontaneously in the crystallization drop to form ADP before crystals formed. However, an unexplained tetrahedrally shaped piece of density was observed close to the position where the γ-phosphate was expected (Figure 2). We modeled this density as a sulfate ion rather than a phosphate ion, because sulfate, being at a concentration of 0.1–0.2 M in the crystallization buffer, would have displaced the cleaved γ-phosphate. This sulfate ion forms numerous contacts through its oxygen atoms with both Mg^{2+} ions, the main-chain amide of the glycine moiety of GSH, the main-chain nitrogen of Gly369 which is located in the
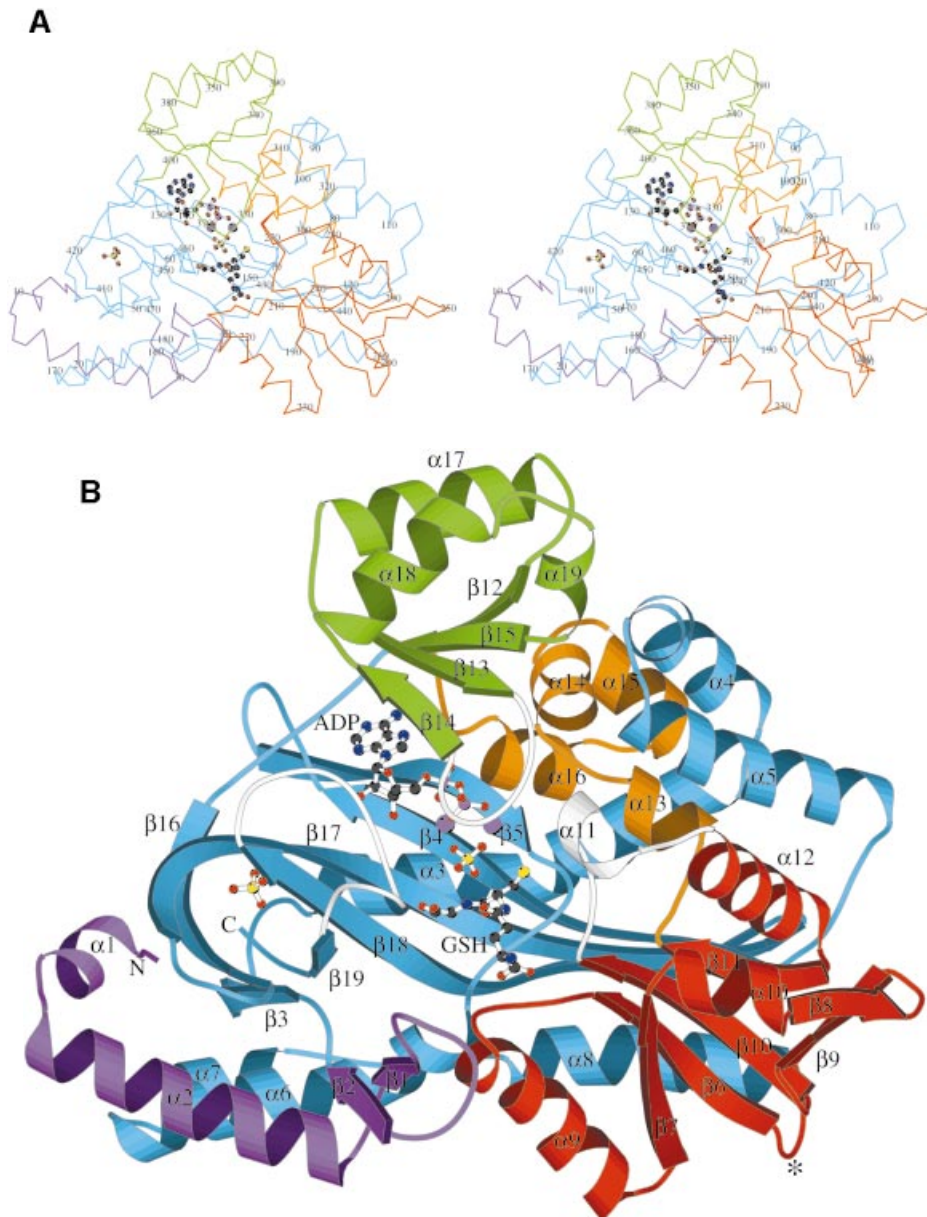
A

B

Fig. 1. Structure of human glutathione synthetase. (**A**) A stereo diagram of a C$_\alpha$ trace of the structure with every tenth residue labeled. (**B**) A ribbon representation of the monomer, shown in the same orientation as in (A), indicating the location of secondary structure. In (A) and (B), ADP, GSH and sulfate ions are shown as ball-and-stick and magnesium ions as mauve spheres. Major structural units (see text) are shown in different colours. The asterisk denotes where the N- and C-termini of the structurally related ecGS are located. These figures were drawn with MOLSCRIPT (Kraulis, 1991).

Gly-rich loop, and two arginine residues (Arg125 and Arg450). A second sulfate ion was located at a crystallographic contact site and forms interactions between symmetry-related molecules.

GSH binding
The GSH molecule is bound at one edge of the parallel β-sheet (Figure 1B). It forms extensive interactions with the protein, including two salt bridges, 11 hydrogen bonding interactions and 82 van der Waals contacts (Figure 2). The γ-glutamyl moiety, through its carboxylate oxygens, forms a salt bridge with Arg267 and hydrogen bonding interactions with the side chains of Ser151, Asn216, Gln220 and the N$_\epsilon$ atom of Arg267. In addition,

Gln220 appears to be fixing Glu214 in an optimum position so it can form a hydrogen bond to the amide nitrogen of the γ-glutamyl moiety. The aromatic side chain of Tyr270 forms a hydrophobic face against the thiol moiety of GSH. Presumably, this environment provides protection against undesirable side-reactions involving the very reactive thiol. The main-chain oxygen of the GSH cysteinyl moiety is in position to form a hydrogen bond with the amide nitrogen of Ser151 and the side chain of Arg125. The amide group of the cysteinyl moiety interacts with the main-chain carbonyl of Ser149. The amide group of the GSH glycyl moiety is within hydrogen bonding distance (2.7 and 2.9 Å) of the two oxygens from the active site sulfate ion. The carboxylate oxygens of the
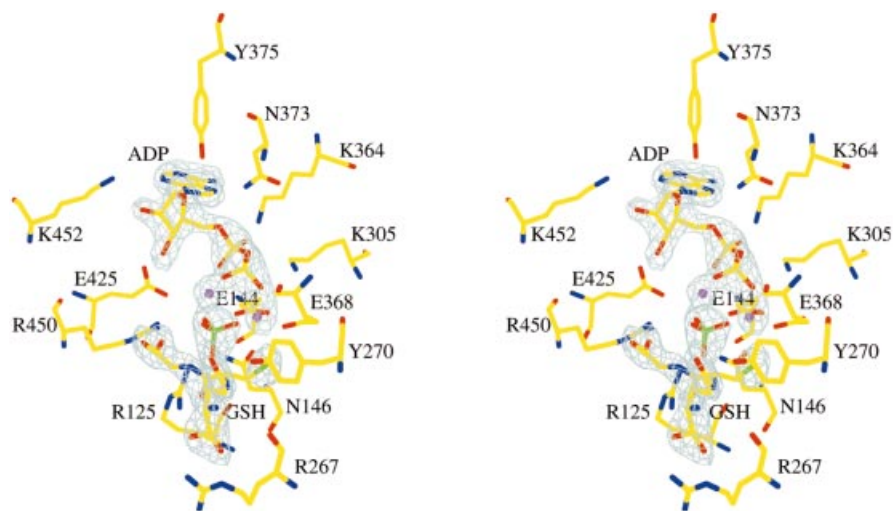
Fig. 2. Difference Fourier (F_o–F_c) electron density map (gray) of the bound ligands at 2.1 Å resolution. The map was calculated from the final model after the bound ligands were omitted from it and a round of positional refinement performed to reduce model bias. The map is contoured at 3.5σ, with the final model overlaid upon it.

glycyl moiety can form hydrogen bonds with the main-chain amides of Val461 and Ala462, both located in the Ala-rich loop between β18 and β19, and also with the side chain of Arg450. The binding pocket around the glycyl moiety is quite tight, in keeping with the probable strict specificity of the enzyme for glycine in this position.

Dimerization

hGS exists as a dimer in solution and the functional dimer in the crystals is generated through a crystallographic 2-fold axis (Figure 3). The intersubunit contacts are extensive (with a total buried surface area of 2702 Å^2) and intimate. The interface is characterized by an elliptically shaped core of hydrophobic residues surrounded by a rim of polar contacts. At the centre of the interface an anti-parallel β-sheet is formed from strands β1 and β2 of one monomer and the corresponding strands in the other monomer. As well as the main-chain interactions between the strands of each monomer, there are hydrogen bonding interactions between the side chains of Tyr47 and Glu43. Helix α9 of one monomer packs against helix α2 of the other monomer at an angle of ~90°. Interactions between the helices include a hydrogen bond between the N_ε atom of Arg221 and Asp24, a hydrophobic interaction between Phe218 and Leu27, and a number of water-mediated contacts. Helix α9 also interacts with helix α6 of the other monomer via Asn231 which forms a water-mediated contact with the main-chain oxygen of Ser168. In addition, there are a number of van der Waals contacts between the two helices. The dimerization unit, consisting of the first two N-terminal strands and helices, is kept in position by the mini-β barrel referred to above (Figure 1B). The dimer interface of hGS is located far away from the active site (~47 Å) and there is no evidence that the active sites act in a dependent fashion of each other. Furthermore, the residues involved in dimerization are not conserved between the sequences of eukaryotic GS enzymes. Given the extensive, intimate contacts between the monomers it appears likely that the dimer confers considerable stability on the human enzyme.
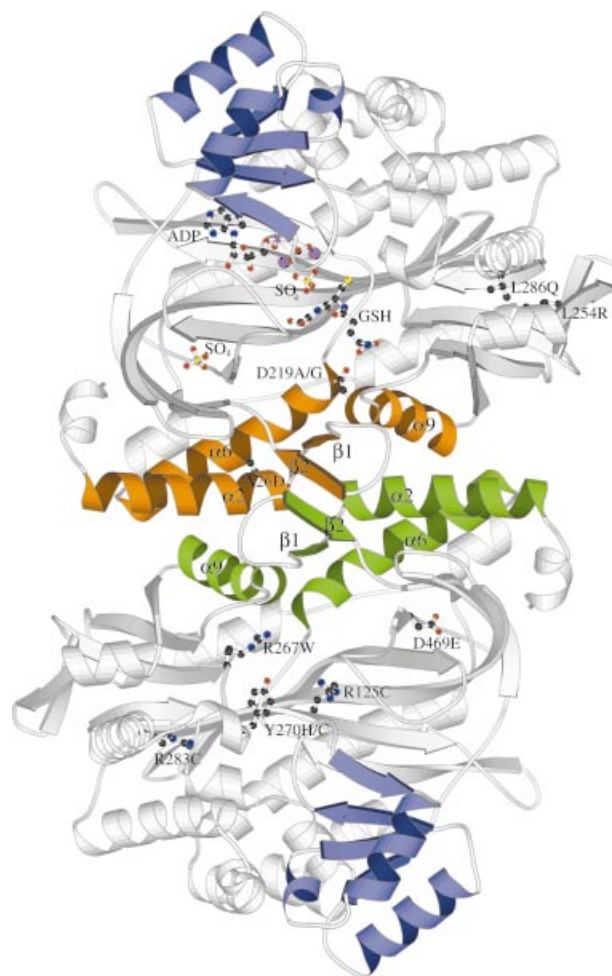

Fig. 3. A ribbon representation of the dimer. The dimerization unit of each monomer is shown in different colour and the lid domains shown in mauve. The GS deficiency mutations are shown in ball-and-stick. This figure was drawn with MOLSCRIPT (Kraulis, 1991).

Discussion

The reaction mechanism

The peptide ligating reaction is thought to proceed in two steps by analogy with other ligases (Artymiuk *et al.*, 1996; Hara *et al.*, 1996). First, the C-terminal carboxylate of γ-glutamylcysteine is phosphorylated by the γ-phosphate group of ATP to form an acylphosphate intermediate. Then, nucleophilic attack by glycine on the acylphosphate intermediate leads to the formation of a tetrahedral carbon intermediate which dissociates to form the product GSH and causes the release of inorganic phosphate and ADP.

The hGS crystal structure, which represents the enzyme–product complex, is fully consistent with the proposed mechanism. The residues that form polar interactions with ATP, Mg^{2+} and GSH in the hGS are strictly conserved in all the sequenced eukaryotic GS enzymes, highlighting the importance of all these contacts. The bound sulfate, mimicking the cleaved γ-phosphate, is located in an approximate line between the β-phosphate of ADP and the amide carbon of the GSH Cys–Gly linkage (Figure 2). This represents an optimal arrangement for generation of the acylphosphate intermediate. The main-chain amide of Gly369 is in position to stabilize the pentavalent phosphate intermediate in the phosphorylation step. The main-chain amide of Ser151 could assist in the second step by stabilizing the charge on the tetrahedral adduct. We have already noted that Arg125 and Arg450 are in a position to neutralize the negative charges of the γ-phosphate. We believe that the main-chain amides play the primary role in the reaction mechanism rather than the arginine residues because they are not strictly conserved in other members of the ATP-grasp superfamily.

It is essential that hGS blocks the active site from the intrusion of solvent during catalysis so as to protect the phosphate intermediate from hydrolytic decomposition. On the other hand, ready access to the active site must be available for the entry of cofactors and substrates and for the exit of products. The enzyme probably accomplishes
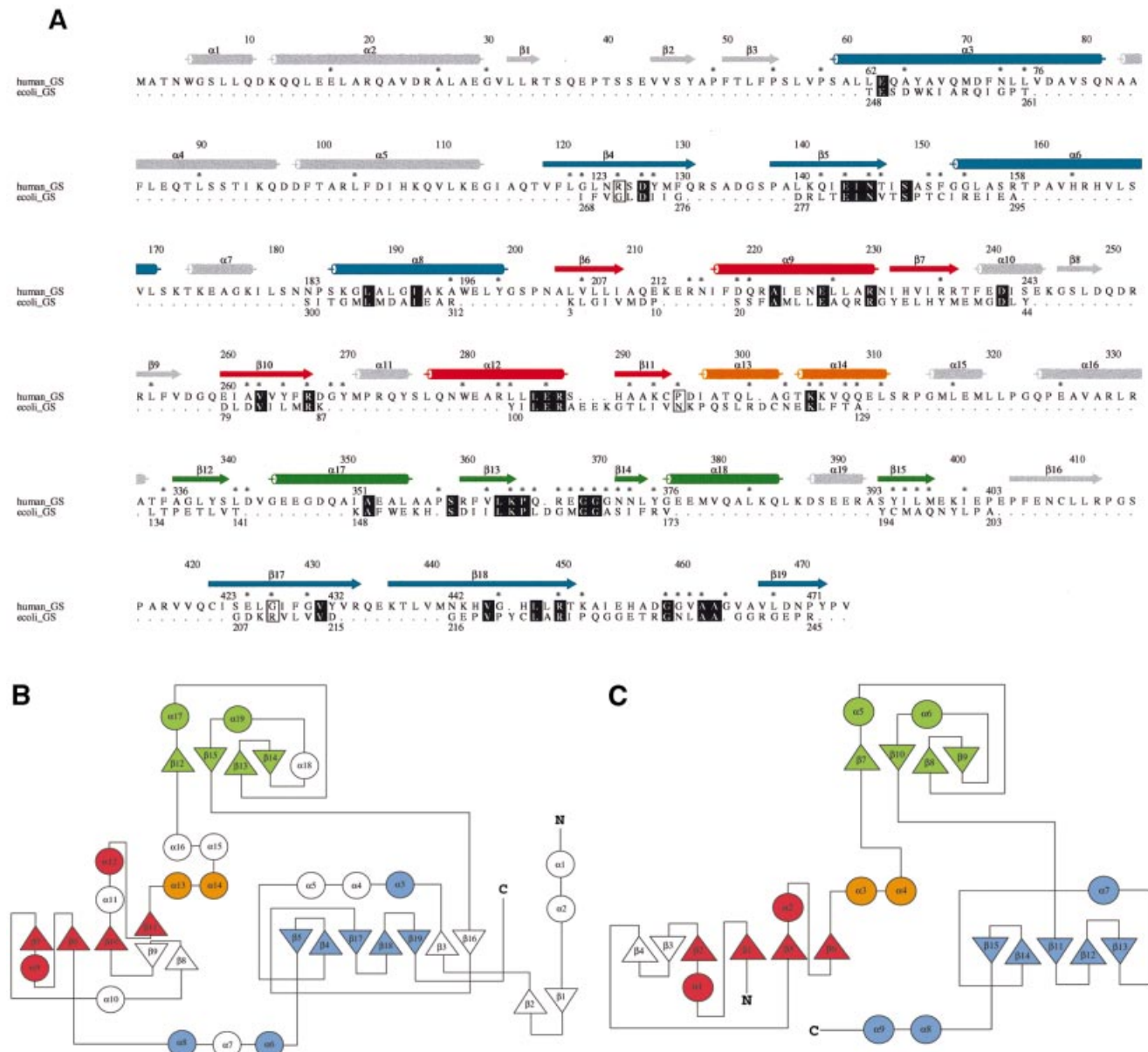
these tasks through the use of the lid domain and the active site loops. The lid domain forms one of the walls of the ATP-binding site (Figure 1B). It is possible that this domain moves during the reaction cycle: in BC the lid domain is located away from the core of the molecule and it has been suggested that the lid would move in to cap the active site in the presence of substrates (Waldrop *et al.*, 1994). Two loops, the Gly-rich and the Ala-rich loop, provide a cover over the active site cleft (Figure 1B) and hence are candidates for movements during the catalytic cycle. Because the loops form many contacts with cofactor and substrate, the loops must move to allow movement of ligands inside and outside of the active site.

The equivalent of these loops are disordered in most structures of the ecGS enzyme (Yamaguchi *et al.*, 1993; Hara *et al.*, 1996), indicating the potential importance of loop movement. The relative importance of lid domain versus loop movement maybe associated with the quaternary associations of the ligase (see below): in ecGS (Yamaguchi *et al.*, 1993) the lid domain is involved in quaternary contacts whereas it is not in hGS or BC (Waldrop *et al.*, 1994). Two aromatic residues, Phe152 and Tyr270, might also play a role in preventing hydrolysis of the intermediate since both residues form hydrophobic walls about the GSH-binding pocket.

A cysteine residue at position 422 has previously been

Table I. Structural effects of mutations in patients with glutathione synthethase deficiency

| Mutation	Structural effect	Recombinant protein activity	Reference
Arg125Cys	γ-glutamyl cysteine binding	⇓	Shi *et al.* (1996)
Asp219Gly	γ-glutamyl cysteine binding	⇓	Shi *et al.* (1996); Dahl *et al.* (1996)
Asp219Ala	γ-glutamyl cysteine binding	NS	Shi *et al.* (1996)
Leu254Arg	γ-glutamyl cysteine binding	NS	Shi *et al.* (1996)
Arg267Trp	γ-glutamyl cysteine binding	⇓	Shi *et al.* (1996); Dahl *et al.* (1996)
Tyr270Cys	γ-glutamyl cysteine binding	⇓	Dahl *et al.* (1996)
Tyr270His	γ-glutamyl cysteine binding	⇓	Dahl *et al.* (1996)
Leu286Gln	γ-glutamyl cysteine binding	NS	Dahl *et al.* (1996)
Asp469Glu	γ-glutamyl cysteine binding	NS	Dahl *et al.* (1996)
Gly464Val	glycine binding	NS	Dahl *et al.* (1996)
Leu188Pro	ATP binding	⇓	Dahl *et al.* (1996)
Ala26Asp	disrupts dimer	⇓	Dahl *et al.* (1996)
Val380 Glu381 deletion	misfolding	insoluble	Shi *et al.* (1996)
Pro314Leu	none obvious, polymorphism?	normal	Shi *et al.* (1996)
Arg283Cys	disulfide formation	⇓	Shi *et al.* (1996); Dahl *et al.* (1996)
Arg330Cys	none obvious, polymorphism?	NS	Dahl *et al.* (1996)

⇓, decrease in activity. NS, not studied.
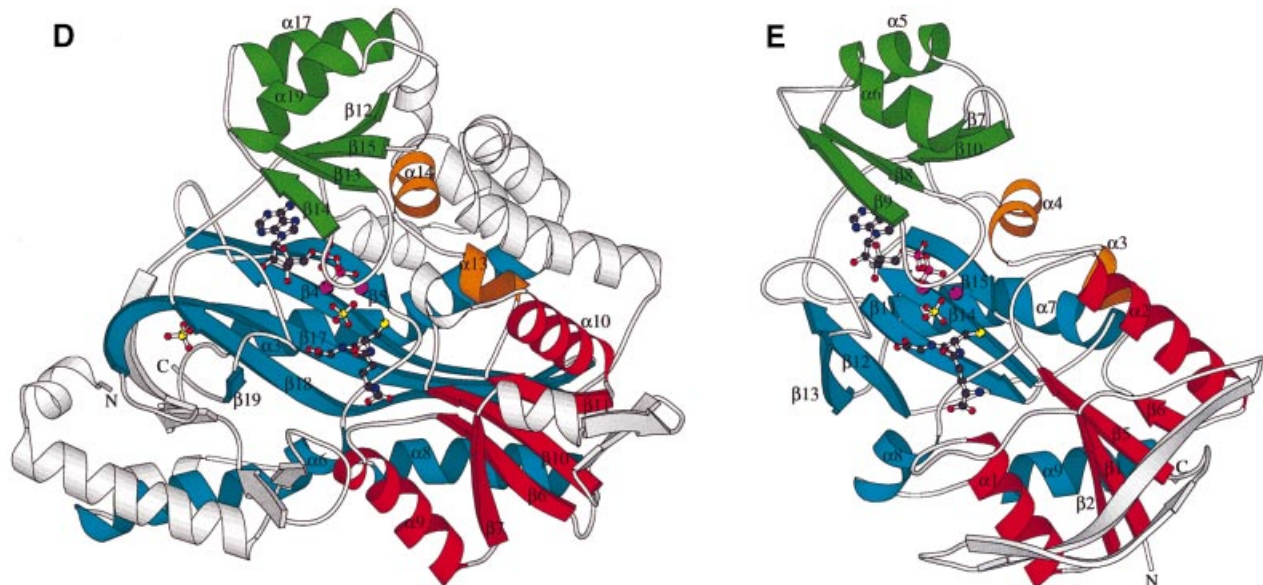
Fig. 4. Comparison of human and bacterial GS. (**A**) Structure-based sequence alignment of hGS and ecGS. The secondary structure and residue numbering of hGS are shown above the alignment and the ecGS numbering below it. The secondary structure of the conserved structural core elements of the ATP-grasp superfamily are coloured as follows: the N-terminal domain is red, the middle, lid domain is green, the C-terminal domain is blue and the linker region (to the lid domain) in orange. Additional secondary structures found in hGS are shown in gray. Strictly conserved residues are highlighted in darkened boxes and invariant residues of GS enzymes from eukaryotic organisms are designated by asterisks. The complimentary mutations and *cis*-peptide regions discussed in the text are highlighted in the open boxes. The figure was produced using ALSCRIPT (Barton, 1993). (**B** and **C**) Topological diagrams highlighting the major structural elements of hGS and ecGS. The colours refer to the structural cores conserved throughout the ATP-grasp superfamily. (**D** and **E**) Ribbon diagrams of hGS and ecGS using the same colour coding as in (A). The ribbon figures were drawn with MOLSCRIPT (Kraulis, 1991).

identified by mutagenesis as being critical for activity (Gali and Board, 1997). This residue is located in the middle of the mini-β barrel (Figure 1B) of which one wall contributes residues to the ATP-binding site. The residue's location suggests that it plays a structural role.

GS deficiency mutations

Several mutations in the GS genes of patients with GS deficiency have been identified, and expression studies on some mutants were performed (Shi *et al.*, 1996; Dahl *et al.*, 1997). Although some homozygous patients were identified, most patients were compound heterozygotes. In two cases, three mutations were observed and in several patients only one mutation was detected in the heterozygous state. The latter group of patients presumably have another allele with a mutation outside the coding region that affects splicing or gene regulation. In all cases, patients had measurable but low GS activity; since the disease is a deficiency and not a total abolition of GS activity, the failure of both GS alleles would probably be lethal.

Since there are deficiency alleles with mutations outside the coding sequence and there are individuals with three mutations, it is not possible to assume that all coding region mutations are responsible for the disease. To help resolve this question the mutations detected in GS-deficient patients have been mapped onto the GS structure with the aim of understanding the molecular basis of GS deficiency (Figure 3; Table I). Many of the mutations appear to affect ligand binding or catalysis. The heterozygous single mutations, Arg125Cys, Asp219Gly, Asp219Ala, Leu254-Arg, Arg267Trp, Try270Cys, Tyr270His, Leu286Gln and Asp469Glu, would all probably interfere with ATP or γ-glutamylcysteine binding. The positive charge of the guanidine group of Arg125 neutralizes the negative charge of the phosphate groups of ATP during the reaction. Not surprisingly, no GS activity has been detected for the Arg125Cys mutant. The Asp219Gly mutation is associated with the mild form of GS deficiency and the mutant protein exhibited substantial residual activity (55% of the wild type) (Shi *et al.*, 1996). Asp219 is engaged in a complicated hydrogen bonding network with active site residues. The residue appears to play an important structural role in the active site but does not itself appear to participate directly in catalysis. Leu254 and Leu286 are constituents of a hydrophobic pocket formed by the packing of helix α12 against the parallel β-sheet and their replacements by polar residues would disrupt the pocket and probably impair the binding of substrate. The Arg267Trp mutation would probably abolish substrate binding because the tryptophan residue is too bulky for the active site. In agreement with this prediction, the mutant protein was found to be devoid of activity (Shi *et al.*, 1996). The phenyl ring of Tyr270 provides a protective environment for the reactive GSH thiol and is probably involved in correct substrate orientation. The cysteine mutation of this residue could result in the formation of a mixed disulfide and inactivate the enzyme. Expression studies of the Tyr270 mutants demonstrated an ~10-fold decrease in enzymic activity compared with the wild type (Dahl *et al.*, 1997). The conservatively substituted mutation, Asp469Glu, is located close to the loop involved in substrate binding. The side chain of

Asp469 can form hydrogen bonds with the side chains of Thr451 and Ser55, as well as the main-chain amides of Asn470 and Ser55. Therefore, the replacement of Asp469 to Glu would cause some adjustments that could result in disturbance of the substrate-binding loop. The Gly464Val mutation would probably alter the conformation and flexibility of the Ala-rich loop involved in binding the glycyl moiety of the substrate. The homozygous single mutation, Leu188Pro, has been shown to seriously impair GS activity (Dahl *et al.*, 1997). This residue is located in helix α8 which packs against strands β4, β17 and β18 of the anti-parallel β-sheet that in turn is involved in ATP binding. Hence the mutation is likely to have a deleterious effect.

Other mutations are likely either to affect dimerization or to disrupt folding. The homozygous Ala26Asp mutation is likely to disrupt the functional dimer of GS because this residue forms part of a hydrophobic pocket neighboring the dimerization interface. Expression of a mutant protein based on the heterozygous mutation Pro314Leu and a deletion missing Val380 and Gln381 resulted in the production of insoluble protein. However, expression of the Pro314Leu mutation alone generated a protein with normal activity (Shi *et al.*, 1996). It is therefore possible that the Pro314Leu mutation is a functionally normal polymorphism. Val380 and Gln381 are located on helix α18, which is a part of the lid domain, and Pro314 forms part of the connection between the parallel β-sheet and the lid domain. The Arg283Cys mutant has been expressed and shown to possess substantially decreased GS activity compared with the wild type (Dahl *et al.*, 1997). We postulate that this behavior might be due to possible disulfide bond formation with Cys294, which is located nearby, during the folding process. The Arg330Cys mutation was found in a patient with the Asp219Gly and Leu286Gln mutations, and given its position on the outside of helix 16 it may be a polymorphism with normal function.

Comparison with other ATP-grasp proteins

The crystal structure of hGS reveals that it does belong to the ATP-grasp superfamily despite the lack of any significant sequence identity with other members. Since extensive structural comparisons between members have been published previously (Fan *et al.*, 1994; Waldrop *et al.*, 1994; Wolodko *et al.*, 1994; Thoden *et al.*, 1997; Esser *et al.*, 1998), we have chosen to restrict our comparisons mainly to the prototype of the family, ecGS (Yamaguchi *et al.*, 1993). Whilst the three common domains are clearly delineated in the existing members of this grouping, that is not true for hGS since additional structural elements have obscured many of the boundaries. For this reason, we have chosen to compare the conserved structural units of each domain: the parallel β-sheet (four-stranded central sheet with pairs of helices packed on both sides, shown in red in Figure 1B), the anti-parallel β-sheet (five-stranded central sheet and three helices, shown in blue in Figure 1B) and the lid domain (four-stranded sheet anti-parallel β-sheet with a pair of helices packed on one side, shown in green in Figure 1B).

The structure of hGS can be superimposed on ecGS with a r.m.s. deviation on C_α atoms of 2.3 Å (for superimposed 204 C_α atoms). The most structurally similar units between the family members are the lid domain and

the anti-parallel sheet unit (Figure 4). The anti-parallel sheet unit has extra helices in hGS, two of which are involved in interactions with the lid domain and its helical connector (α4 and α5) and another helix (α7) together with an extended helix α6 which are involved in orienting the dimerization unit (Figures 1B and 3). The parallel β-sheet unit superimposes closely with ecGS although deviations occur in the helices that pack onto the sheet. The helical connection between the parallel β-sheet unit and the lid domain consists of two short helices (α13 and α14) that have the same lengths and orientations in ecGS. In hGS there are two further short helices (α15 and α16) in the connection. A rare, non-proline, *cis*-peptide bond in the helical connection has been noted previously in ecGS (preceding Asn114; Yamaguchi *et al.*, 1993) and synapsin (preceding Asn214; Esser *et al.*, 1998). Interestingly, the corresponding residue in BC (Waldrop *et al.*, 1994) and DDL (Fan *et al.*, 1994) is glycine, and a *cis*-proline (Pro295) in hGS. There is a considerable degree of variation in the amino acids forming the ATP-binding pocket (Figure 4A). Only three residues are strictly conserved in virtually all ATP-grasp enzymes: Lys305 (Arg in CPS), Lys364 (Arg in CPS) and Glu144. The lysine residues interact with the ATP phosphate groups and the glutamate is involved in magnesium binding.

The quaternary structures of family members differ significantly. The most similar quaternary organization to hGS is found in BC (Waldrop *et al.*, 1994) where the dimerization unit is found in the same spatial location as the former although the unit is located at the C-terminal end rather than the N-terminal end of the molecule. The ecGS molecule is a tetramer but can be considered as a loose dimer of tight dimers. Both types of dimer interaction involve interactions with the lid domain.

Gene permutation

Previous workers have argued that members of the ATP-grasp superfamily possess a common evolutionary ancestor (Fan *et al.*, 1995; Artymiuk *et al.*, 1996). Structure-based alignments of three members (ecGS, DDL and BC) revealed three conserved sequence motifs (Artymiuk *et al.*, 1996). The first motif (hGS residues 351–376) corresponds to part of the lid domain and the Gly-rich loop. The second (hGS residues 394–403, 424–431) and the third (hGS residues 122–130, 142–148) motifs form some of the walls of the ATP-binding site. A search of sequence databases with these motifs revealed a total of 15 members of the superfamily but most interestingly this search failed to identify hGS as a member (Galperin and Koonin, 1997). The crystal structure reported here reveals that hGS shares close structural similarities with the family.

The most surprising feature of the hGS structure is that its core structural units have been circularly permuted with respect to all the other family members as shown in Figure 4: the ATP-grasp C-terminal domain is split into two in hGS so that approximately half is located at the N-terminus and the other half is located at the C-terminus. This permutation is particularly surprising given the close functional similarities between two members, hGS and ecGS, where the common domain folds and the location of the ATP- and GSH-binding sites are well preserved. The N- and C-termini of each enzyme are adjacent (Figures 1B and 4D, E) supporting the view that the

circular permutations are reversible at the genetic level. The pairwise sequence identity between hGS and ecGS after structure-based sequence alignment is only 10% (Figure 4A). Analysis of the conserved ATP-grasp sequence motifs (Artymiuk *et al.*, 1996) indicates that these two enzymes are no more closely related to each other than to other family members (data not shown). Furthermore, only two out of a total of 11 residues involved in side-chain interactions with GSH are strictly conserved between the two enzymes. These observations, taken together with the different modes of oligomerization, provide the basis for arguing that hGS did not evolve from ecGS but rather both enzymes evolved from a common, more distant, ancestor. Furthermore, the domain containing the parallel β-sheet unit in hGS most closely resembles the equivalent domain in ecGS and synapsin and, in particular, the β8–β9 hairpin is unique to these structures. These observations suggest that the last common ancestor of the three diverged from the other family members before the circular permutation event. The human enzyme is highly similar to other eukaryotic GS enzymes with the lowest pairwise sequence identity of 35% with a yeast GS. All residues involved in polar interactions with ligands in the hGS structure are either strictly conserved or conservatively substituted in the other eukaryotic GS enzymes (Figure 4A; Whitbread *et al.*, 1998). These observations show that eukaryotic GS enzymes form a closely related family which is very different from their bacterial counterparts.

The circular permutation seen here most probably arose by tandem duplication of a single ancestral gene encoding all three domains followed by deletions at both ends of the gene to yield the direct ancestor of hGS. Such a mechanism has been proposed for aconitase evolution (Gruer *et al.*, 1997). Consistent with this fusion hypothesis, is the location of an exon boundary after residue 204 (the location of the N-terminus of the ATP-grasp N-terminal domain). There are very few known examples of naturally occurring circularly permuted proteins and most of these have been identified by sequence comparisons before structures became available (Lindqvist and Schneider, 1997), which was not possible in this case. A remarkable feature of the genetic permutation is how the permuted secondary structure combines, irrespective of order, to form a very similar active site located at the domain boundaries. Future sequence-based searches for new members of the ATP-grasp superfamily will need to allow for possible gene permutations.

Materials and methods

Crystallization and data collection

hGS was expressed and purified as described previously (Gali and Board, 1997). The protein was dialysed against 10 mM Tris–HCl pH 7.5, 10 mM MgSO$_4$, 1 mM dithiothreitol and 0.1 mM ATP or ADP and concentrated to between 3 and 5 mg/ml prior to the crystallization. The protein was mixed with an equivolume ratio of a solution containing 100 mM MES pH 6.0, 10 mM GSH, 5 mM EDTA, 0.1–0.2 M MgSO$_4$, 9–12% PEG 4K and equilibrated over the same solution using the hanging drop vapour diffusion technique at 22°C. The crystals grew to a size of up to ~0.2×0.2×0.4 mm within a week. They belong to the space group P4$_1$22 with cell dimensions $a = b = 84.3$ Å, $c = 197.6$ Å and contain one monomer per asymmetric unit with a solvent content of ~60%.

Prior to soaking the crystals in heavy atom solutions, GSH had to be

Table II. Structure determination statistics

Diffraction data statistics

Data set	No. of crystals used	Resolution (Å)	Completeness (%)	I/σ in the last resolution shell	R_{merge} (%)[a]	No. of measurements	No. of unique reflections
Native RT	1	20–3.0	96.1 (61.6)	5.7	12.3 (32.0)	169 155	15 198
Cis-Pt$(NH_3)_2Cl_2$	2	20–3.2	99.6 (98.2)	4.5	12.7 (28.2)	161 664	13 472
Pt(ethylenediamine)Cl_2	1	20–3.2	96.4 (99.2)	1.5	23.4 (57.6)	43 178	13 344
Hg(II)Cl_2	1	20–3.1	94.4 (96.1)	1.7	20.0 (78.2)	69 243	13 362
$KPtCl_6$	1	20–3.3	93.5 (41.0)	4.0	12.8 (29.1)	60 816	10 996
Native cryo I	1	20–2.4	92.4 (68.8)	1.5	11.2 (45.2)	185 473	26 951
Native cryo II	1	15–2.1	95.9 (95.9)	1.7	5.7 (19.7)	201 501	40 495

Phasing statistics

Derivative	Soaking time	Binding sites	Resolution (Å)	R_{iso} (%)[b]	Phasing power[c]		R_{cullis}[d]	
					Centric	Acentric	Centric	Acentric
cis-Pt$(NH_3)_2Cl_2$	2 days	Met316, Met271, Met70, His290	20–3.2	31.3	0.88	1.1	0.78	0.83
Pt(ethylenediamine)Cl_2	4 days	Met316, Met271, Met70	20–3.2	29.7	0.92	1.14	0.76	0.83
Hg(II)Cl_2	45 min	Cys294, Cys409, His233, Cys422, His444	20–3.1	34.2	0.84	1.08	0.82	0.86
$KPtCl_6$	1 day	Met271, His444	20–5.0	20.3	0.48	0.63	0.93	0.92

Final refinement statistics

Data set	Native native cryo I and cryo II were merged using DATAMAN (Jones *et al.*, 1991) with R_{merge} of 11.0%
Resolution	20–2.1 Å
R-factor[e]/R_{free}[f]	21.8/27.8%
R.m.s.d. bond lengths	0.011 Å
R.m.s.d. bond angles	2.1°
R.m.s.d. dihedral angles	19.8°
R.m.s.d. bonded Bs (m–c)	4.0 Å^2
R.m.s.d. bonded Bs (s–c)	3.8 Å^2

Values given in parentheses are for the highest resolution shell (~0.1 Å in width).
[a]$R_{merge} = \Sigma (I-\langle I \rangle)/\Sigma(I)$, where I is the intensity measurement for a given reflection, $\langle I \rangle$ is the average intensity for multiple measurements of this reflection.
[b]$R_{iso} = \Sigma ||F_{PH}|-|F_P||/\Sigma|F_P|$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively.
[c]Phasing power = root-mean-square of $(|F_H|/E)$, where F_H is the calculated heavy atom structure factor amplitude and E is the residual lack-of-closure error.
[d]$R_{cullis} = \Sigma|||F_{PH}| \pm |F_P||-|F_H|/\Sigma||F_{PH}| \pm |F_P||$.
[e]R-factor $= \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[f]R_{free} was calculated with 5% of the diffraction data that were selected randomly and not used throughout refinement.

removed from the solution because of its reactive thiol and at the same time the concentration of $MgSO_4$ had to be lowered to 0.1 M, otherwise the crystals would dissolve. The details of the heavy atom soaks are given in Table II. The following heavy atoms proved only slightly soluble in artificial mother liquor and were hence prepared as saturated solutions: *cis*-Pt(diamine)dichloride, Pt(ethylenediamine)dichloride and $HgCl_2$. $KPtCl_6$ was made up at a concentration of 1 mM. The 'Native RT' data set and all derivative data sets were collected at room temperature on an in-house MARResearch imaging (180 mm diameter) plate detector. The resolution of these data sets were limited to 3 Å because of the long c axis. The 'Native cryo I' data set was collected in-house from a single crystal frozen at 100 K using a MARResearch imaging (345 mm diameter) plate. In this case, 15% MPD was used as a cryoprotectant. The 'Native cryo II' data set was collected from a single crystal frozen to 100 K using synchrotron radiation at the BioCARS beamline, 14-BM-C, Advanced Photon Source, Chicago, IL. The in-house data were processed using the HKL (Otwinowski and Minor, 1997) and CCP4 software packages (CCP4, 1994). The synchrotron data were collected on both a MARResearch image plate scanner and an ADSC Quantum-4 CCD detector and processed using either the HKL (Otwinowski and Minor, 1997) or the MOSFLM package (Leslie, 1992). Subsequently, it proved beneficial to merge the two native data sets

collected from frozen crystals because the low resolution data from the in-house data set were more reliably measured.

Structure determination, model building and refinement
The locations of heavy atom sites for the *cis*-Pt(diamine)dichloride derivative were determined using the programs RSPS and VECREF of the CCP4 package (CCP4, 1994). The positions of other heavy atom derivatives were found from difference Fourier maps. The heavy atom positions for all four derivatives were refined and the phases calculated using the program MLPHARE (CCP4, 1994). The overall figure-of-merit (FOM) was 0.35 for all data to 3.0 Å resolution (with a FOM of 0.50 to 4.4 Å resolution). The phases were improved by density modification with DM (CCP4, 1994) resulting in a new FOM of 0.59. The first electron density map calculated from these phases was readily interpretable. An initial polyalanine model could be constructed, with the aid of skeletonized density maps and the program O (Jones *et al.*, 1991), which constituted ~50% of all C_α atoms. Several cycles of refinement with REFMAC (Murshudov *et al.*, 1997), making use of the density-modified phases and structure factors from the Native RT data set, interspersed with rounds of model building, resulted in a polyalanine model consisting of ~70% of all C_α atoms. The first region of sequence was identified from the positions of two mercury-

binding sites, which were located in close proximity to two C$_\alpha$ positions separated by 12 residues. This lead to the unambiguous identification of Cys409 and Cys422. Further rounds of refinement were performed incorporating the amino acid sequence as it was identified and the phase restraints were dropped. Upon convergence, the model included residues 14–131 and 140–474, one molecule each of ADP and GSH, two Mg^{2+} ions and two sulfate ions. The R-factor and R_{free} were, respectively, 22.3 and 28.7% to a resolution of 3.0 Å. The refinement was then continued with the Native cryo II data set. Because the cell parameters for the frozen crystal data sets were slightly different from that of data collected at room temperature, a round of rigid body refinement was necessary. Several cycles of refinement with REFMAC (Murshudov *et al.*, 1997), interspersed with rounds of model building resulted in a final model of hGS consisting of residues 3–474, one molecule each of ADP and GSH, two Mg^{2+} ions, two sulfate ions and 230 water molecules. The refinement statistics are given in Table II. The final model is of good quality with 92% of residues in the most favorable regions of the Ramachandran plot (Laskowski *et al.*, 1993). Only Glu436 falls in a disallowed region: this residue is located in the second position of a type IV β-turn and forms interactions with Ser118 of a neighboring strand. The quality of the model is supported by the excellent electron density throughout the molecule (see Figure 2). The correctness of the model was confirmed by the convincing fit of all 40 aromatic residues, the location of heavy atom sites in chemically reasonable positions (Table II) and 3D-1D scores (Lüthy *et al.*, 1992) that never fell below 0.15. Atomic coordinates have been deposited in the Brookhaven Protein Data Bank (accession code 2HGS) and will be available at the time of publication from either the Data Bank or by e-mail from the authors (e-mail: galina@bardoc.medstv.unimelb.edu.au).

Acknowledgements

References

Anderson,M.E. (1998) Glutathione: an overview of biosynthesis and modulation. *Chem. Biol. Interact.*, **111–112**, 1–14.

Artymiuk,P.J., Poirrette,A.R., Rice,D.W. and Willett,P. (1996) Biotin carboxylase comes into the fold. *Nature Struct. Biol.*, **3**, 128–132.

Barton,G.J. (1993) ALSCRIPT—a tool to format multiple sequence alignments. *Protein Eng.*, **6**, 37–40.

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Dahl,N., Pigg,M., Ristoff,E., Gali,R., Carlsson,B., Mannervik,B., Larsson,A. and Board,P. (1997) Missense mutations in the human glutathione synthetase gene result in severe metabolic acidosis, 5-oxoprolinuria, hemolytic anemia and neurological dysfunction. *Human Mol. Genet.*, **6**, 1147–1152.

Esser,L., Wang,C.-R., Hosaka,M., Smagula,C.S., Südhof,T.C and Deisenhofer,J. (1998) Synapsin I is structurally similar to ATP-utilizing enzymes. *EMBO J.*, **17**, 977–984.

Fan,C., Moews,P.C., Walsh,C.T. and Knox,J.R. (1994) Vancomycin resistance: Structure of D-alanine:D-alanine ligase at 2.3 Å resolution. *Science*, **266**, 439–443.

Fan,C., Moews,P.C., Shi,Y., Walsh,C.T. and Knox,J.R. (1995) A common fold for peptide synthetases cleaving ATP to ADP: glutathione synthetase and D-alanine:D-alanine ligase of *Escherichia coli. Proc. Natl Acad. Sci. USA*, **92**, 1172–1176.

Gali,R.R. and Board,P.G. (1995) Sequencing and expression of a cDNA for human glutathione synthetase. *Biochem. J.*, **310**, 353–358.

Gali,R.R. and Board,P.G. (1997) Identification of an essential cysteine residue in human glutathione synthase. *Biochem. J.*, **321**, 207–210.

Galperin,M.Y. and Koonin,E.V. (1997) A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci.*, **6**, 2639–2643.

Gruer,M.J., Artymiuk,P.J. and Guest,J.R. (1997) The aconitase family: three structural variations on a common theme. *Trends Biochem. Sci.*, **22**, 3–6.

Gushima,H., Yasuda,S., Soeda,E., Yokota,M. and Kimura,A. (1984) Complete nucleotide sequence of the *E.coli* glutathione synthetase gsh-II. *Nucleic Acids Res.*, **12**, 9299–9307.

Hara,T., Kato,H., Katsube Y. and Oda,J. (1996) A pseudo-Michaelis quaternary complex in the reverse reaction of a ligase: structure of *Escherichia coli* B glutathione synthetase complexed with ADP, glutathione and sulfate at 2.0 Å resolution. *Biochemistry*, **35**, 11967–11974.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of error in these models. *Acta Crystallogr.*, **A42**, 110–119.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of proteins. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Leslie,A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. In Wolf,W.M. and Wilson,K.S. (eds), *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography No. 26.* SERC Laboratory, Daresbury, Warrington, UK.

Lindqvist,Y. and Schneider,G. (1997) Circular permutations of natural protein sequences: structural evidence. *Curr. Opin. Struct. Biol.*, **7**, 422–417.

Lüthy,R., Bowie,J.U. and Eisenberg,D. (1992) Assessment of protein models with 3-dimensional profiles. *Nature*, **356**, 83–85.

Meister,A. and Larsson,A. (1995) Glutathione synthetase deficiency and other disorders of the γ-glutamyl cycle. In Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle D. (eds), *The Metabolic Basis of Inherited Diseases.* 7th edn. Wiley, New York, NY, pp. 1461–1477.

Murzin,A.G. (1996) Structural classification of proteins: new superfamilies. *Curr. Opin. Struct. Biol.*, **6**, 386–394.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **D53**, 240–255.

Otwinowski,Z. and Minor,W. (1997) Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Ristoff,E. and Larsson,A. (1998) Patients with genetics defects in the γ-glutamyl cycle. *Chem. Biol. Interact.*, **111–112**, 113–121.

Shi,Z.Z., Habib,G.M., Rhead,W.J., Gahl,W.A., He,X., Sazer,S. and Lieberman,M.W. (1996) Mutations in the glutathione synthetase gene cause 5-oxoprolinuria. *Nature Genet.*, **14**, 362–365.

Thoden,J.B., Holden,H.M., Wesenberg,G., Raushel,F.M. and Rayment,I. (1997) Structure of carbamoyl phosphate synthetase: A journey of 96 Å from substrate to product. *Biochemistry*, **36**, 6305–6316.

Waldrop,G.L., Rayment,I. and Holden,H.M. (1994) Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry*, **33**, 10249–10256.

Webb,G.C., Vaska,V.L., Gali,R.R., Ford,J.H. and Board,P.G. (1995) The gene encoding human glutathione synthetase (GSS) maps to the long arm of chromosome 20 at band 11.2. *Genomics*, **30**, 617–619.

Whitbread,L., Gali,R.R. and Board,P.G. (1998) The structure of the human glutathione synthetase gene. *Chem. Biol. Interact.*, **111–112**, 35–40.

Wolodko,W.T., Fraser,M.E., James,M.N.G. and Bridger,W.A. (1994) The crystal structure of succinyl-CoA synthetase from *Escherichia coli* at 2.5 Å resolution. *J. Biol. Chem.*, **269**, 10883–10890.

Yamaguchi,H., Kato,H., Hata,Y., Nishioka,T., Kimura,A., Oda,J. and Katsube,Y. (1993) Three-dimensional structure of the glutathione synthetase from *Escherichia coli* B at 2.0 Å resolution. *J. Mol. Biol.*, **229**, 1083–1100.