J. Mol. Biol. (1997) 268, 869–885

JMB

The Crystal Structure of Seleno-Glutathione Peroxidase from Human Plasma at 2.9 Å Resolution

Bin Ren[1], Wenhu Huang[2], Björn Åkesson[2] and Rudolf Ladenstein[1]*

[1]*Karolinska Institute, Novum Center for Structural Biochemistry, S-141 57 Huddinge, Sweden*

[2]*Chemical Center, University of Lund, Department of Applied Nutrition and Food Chemistry S-221 00 Lund, Sweden*

**Corresponding author*

Glutathione peroxidase belongs to the family of selenoproteins and plays an important role in the defense mechanisms of mammals, birds and fish against oxidative damage by catalyzing the reduction of a variety of hydroperoxides, using glutathione as the reducing substrate. However, the physiological role of human plasma glutathione peroxidase remains unclear due to the low levels of reduced glutathione in human plasma and the low reactivity of this enzyme.

The crystal structure of human plasma glutathione peroxidase was determined by Patterson search methods using a polyalanine model modified from the known structure of bovine erythrocyte glutathione peroxidase. The structure was refined to a crystallographic *R*-factor of 0.228 ($R_{free} = 0.335$) with $I > 2\sigma(I)$ reflections in the resolution range of 8 to 2.9 Å. The asymmetric unit contains a dimer. Tetramers are built up from dimers by crystallographic symmetry.

The subunit structure of the plasma enzyme shows the typical structure motif of the thioredoxin fold consisting of a central β-sheet and several flanking α-helices. The active site selenocysteine residue is situated in the loop between β1 and α1 and is located in a pocket on the protein surface. The overall structure of the human plasma enzyme is similar to that of the bovine erythrocyte enzyme. The main differences in their subunit structures are an extended N terminus and the possible existence of a disulfide bridge in the plasma enzyme. Compared to the bovine erythrocyte enzyme, a number of residues in the active site are mutated or deleted in the plasma enzyme, including all the residues that were previously suggested to be involved in glutathione binding. The observed structural differences between the two enzymes suggest differences in substrate binding and specificity.

© 1997 Academic Press Limited

Keywords: glutathione peroxidase; crystal structure; Patterson search; selenoprotein; human plasma

Introduction

Mammalian glutathione peroxidase (GSHPx; EC 1.11.1.9) has been characterized as a cytosolic selenoenzyme that protects biomembranes and other cellular components against oxidative damage (Mills, 1957; Flohé, 1989). It catalyzes the reduction of hydrogen peroxide and a variety of organic hydroperoxides using glutathione as the reducing substrate. The classical cellular enzyme is a tetramer and contains a selenocysteine residue in each of the identical subunits. The selenocysteine residue is encoded by an opal stop codon UGA (Chambers *et al.*, 1986; Zinoni *et al.*, 1986) and is located in the active site of the enzyme (Forstrom *et al.*, 1978b). Selenium deficiency results in a decrease of the enzymatic activity (Reiter & Wendel, 1983). Kinetic studies have shown that the reduction of hydroperoxides by the cellular enzyme follows a ter uni ping-pong mechanism (Flohé *et al.*, 1972). The crystal structure of bovine erythrocyte GSHPx has been determined and crystallographically refined at 2 Å resolution (Ladenstein *et al.*,

Abbreviations used: GSHPx, glutathione peroxidase; pGSHPx, plasma glutathione peroxidase; cGSHPx, cellular glutathione peroxidase; GSH, reduced glutathione; Cso, selenocysteine; NCS, non-crystallographic symmetry; r.m.s., root-mean-square; σ, standard deviation; DTT, dithiothreitol; PEG, polyethylene glycol; MPD, 2-methyl-2,4-pentanediol; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; V_M = (volume of unit cell)/(molecular mass in cell) Å^3/Dalton.

0022–2836/97/200869–17 $25.00/0/mb971005

© 1997 Academic Press Limited

1979; Epp *et al.*, 1983). The structural studies have provided the first view of the active site with its selenocysteine residue, which is involved in catalysis and probably shuttles between a selenolate anion ($R\text{-}Se^{\ominus}$) and a selenenic acid (R-Se-OH) in the catalytic cycle. Substrate binding studies on the bovine cellular enzyme revealed that only two glutathione molecules bind to each tetramer, which indicated an apparent half-site reactivity of the enzyme (Epp *et al.*, 1983). Further crystallographic studies on enzyme-substrate/product interactions were hampered by half occupancy of the active sites.

In addition to the classical cellular enzyme, two other intracellular glutathione peroxidases have been identified in mammals. The gastrointestinal tract GSHPx shares most of its enzymatic properties with the classical cellular enzyme but shows a higher reactivity towards organic hydroperoxides (Chu *et al.*, 1993). The enzyme has been suggested to play a major role in protecting mammals from the toxicity of ingested lipid hydroperoxides. Phospholipid hydroperoxide GSHPx is a membrane-bound enzyme, which exists as monomer and particularly reduces phospholipid and cholesterol hydroperoxides in biological membranes (Thomas *et al.*, 1990; Schuckelt *et al.*, 1991).

Unlike other GSHPxs, human plasma glutathione peroxidase is an extracellular enzyme, which has been purified and characterized by Broderick *et al.* (1987), Takahashi *et al.* (1987) and Maddipati & Marnett (1987). Existing as a tetramer in its native state, the plasma enzyme has been shown to be structurally, enzymatically and antigenically different from other GSHPx enzymes. It has been reported to be a glycoprotein and did not cross-react with antisera against the other enzymes (Takahashi *et al.*, 1987; Chu *et al.*, 1993). Plasma GSHPx shows distinct kinetic properties when reacting with glutathione, hydrogen peroxide, organic hydroperoxides and phospholipid hydroperoxides (Takahashi *et al.*, 1987; Maddipati & Marnett, 1987; Yamamoto & Takahashi, 1993; Esworthy *et al.*, 1993). While the classical cellular enzyme seems to be a constitutive form in all tissues, the plasma enzyme exhibits tissue-specific expression. Human kidney proximal tubules have been shown to be the main source (Avissar *et al.*, 1994). However, plasma enzyme mRNA is present and expressed in human liver, heart, lung, breast, and eye (Chu *et al.*, 1992; Martin-Alonso *et al.*, 1993). In addition to human plasma, 90% of the GSHPx activity in human milk has been attributed to this enzyme (Avissar *et al.*, 1991).

The existence of multiple forms of GSHPx is due to the expression of different genes (Takahashi *et al.*, 1990; Schuckelt *et al.*, 1991; Chu *et al.*, 1993). Within each GSHPx class high sequence identity is observed, while between classes the identity is much lower. The primary structure of human plasma GSHPx shows 44% identity with the human cellular enzyme, 34% with the human gastrointestinal tract enzyme and 24% with the human phospholipid GSHPx (Takahashi *et al.*, 1990; Chu *et al.*, 1993; Esworthy *et al.*, 1994). Here, we report the crystal structure analysis and the crystallographic refinement of human plasma GSHPx.

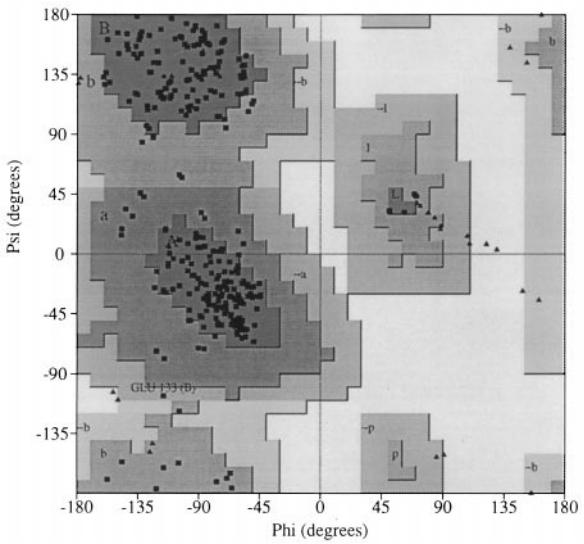

Figure 1. Ramachandran plot of the main-chain dihedral angles for residues in the dimer of human pGSHPx. Glycine residues are represented by triangles and all other residues by squares. The most favored regions (A, B and L) are shaded in dark gray and the disallowed regions are in white. Between them are the additionally allowed regions (a, b, 1 and p) and generously allowed regions (~a, ~b, ~1 and ~p). Of the residues in the pGSHPx dimer, 86.9% are in the most favored regions and one residue, Glu133B, in the generously allowed region. The Figure was produced with PROCHECK (Laskowski *et al.*, 1993).

Results

Quality of the model

An asymmetric unit contains a dimer with two subunits, designated A and B. Each subunit consists of 187 residues with 1502 non-hydrogen atoms. The model was refined with NCS restraints to a crystallographic *R*-factor of 22.8% with 8101 unique reflections ($I_o > 2\sigma(I_o)$) in the resolution range 8.0 to 2.9 Å. The final free *R*-factor was 33.5%. The r.m.s. deviations of the model from ideal geometry were 0.017 Å for bond lengths and 2.1° for bond angles (see Materials and Methods). The Ramachandran plot (Figure 1) of the main-chain conformational angles shows that 86.9% of the residues lie within the most favored regions and 12.8% in additionally allowed regions. Only the residue Glu133B lies in the generously allowed region. This residue is situated in a flexible loop and badly defined in the electron density map. The averaged temperature factor of the refined model for all atoms is 11.8 $Å^2$, which agrees well with the statistically estimated *B*-factor of 13.0 $Å^2$, obtained from a Wilson plot (Wilson, 1949). Because NCS

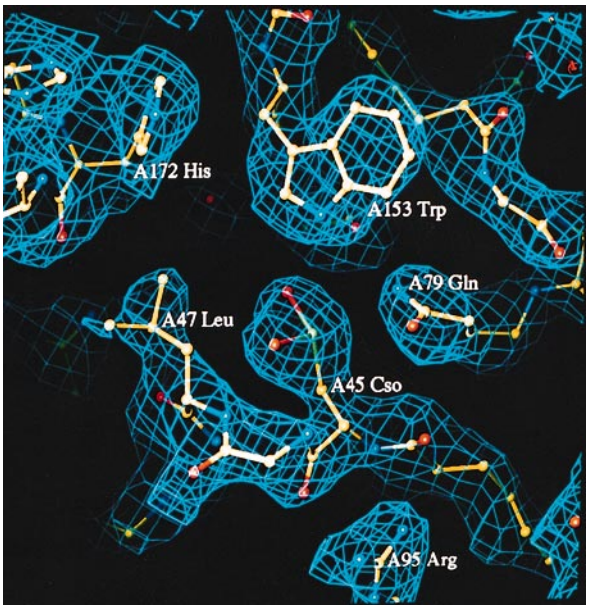

Figure 2. Section of the final $(2F_o - F_c)$ electron density map at the active site region including the selenocysteine residue (Cso45). Carbon atoms are shown in yellow, oxygen atoms in red, nitrogen atoms in blue and selenium in green. The map is contoured at the 1σ level.

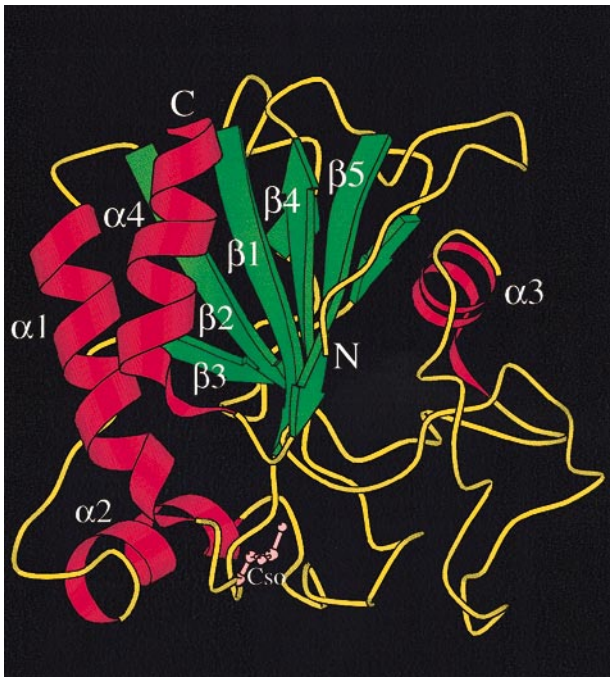

Figure 3. Ribbon diagram of a subunit of human pGSHPx. Helices are shown in red, β-strands in green and other structure elements in yellow. The active site residue, selenocysteine (Cso45), is shown by ball-and-stick representation. The Figure was produced with MOLSCRIPT (Kraulis, 1991).

restraints were applied in the refinement, the two subunits in the asymmetric unit exhibit nearly identical main-chain conformations. The r.m.s. deviations between the aligned residues in the two subunits are 0.18 Å for C^{α} atoms and 0.83 Å for all atoms. The mean error in the atomic coordinates, estimated from a Luzzati plot (Luzzati, 1952), is 0.34 Å.

The main-chain atoms are generally well defined in both $(2F_o - F_c)$ and averaged electron density maps. The $(2F_o - F_c)$ map (Figure 2) shows clear densities for most of the side-chain atoms, whereas some of them are not well defined in the averaged map. These side-chains are usually located in flexible loops on the protein surface or have different conformations in the two subunits. This could be due to their different packing environments in some cases. The breakdown of the non-crystallographic symmetry in these regions may lead to poorly defined electron densities observed in the averaged map. The residues between Gly136 and Asp139 show diffuse densities in both subunits. These residues are located in a flexible loop and constitute part of the active site pocket on the protein surface (see below).

The primary structure of human pGSHPx was deduced from the cDNA sequence, which probably contains a signal peptide at the N terminus (Takahashi *et al.*, 1990). The N-terminal residue and the exact size of the native protein isolated from human plasma is still unknown. During model building, the first residue at the N terminus was traced by a lysine residue, which corresponds to the 29th residue in the sequence. From this lysine residue on, the amino acid residues in each subunit are numbered 1 to 187 in the present model. The additional residues beyond Lys1 and Met187 at both termini could not be identified in the averaged electron density map.

Subunit structure

Glutathione peroxidase from human plasma exists as a tetramer in solution and in the crystal (see below). Each asymmetric unit consists of two identical subunits related by a non-crystallographic 2-fold axis. The subunit structure shows the typical structure motif of the thioredoxin fold (Martin, 1995), which consists mainly of a central β-sheet and several flanking α-helices (Figure 3). The secondary structure elements within a subunit are listed in Table 1.

The central β-sheet is composed of five β-strands. They are arranged in the order β3 (residues 105 to 107), β2 (residues 65 to 71), β1 (residues 35 to 39), β4 (residues 157 to 161) and β5 (residues 165 to 170). Three β-strands close to the N terminus, β1, β2 and β3, are parallel, and the other two close to the C terminus, β4 and β5, are antiparallel. The β-sheet is twisted by about 45° when viewed from the side. The topology of this sheet can be described as 1x, 1x, −3, −1 (Richardson, 1981), where x represents a crossover connection between strands. Aside from the central β-sheet there are four flanking α-helices. Three of

Table 1. Secondary structure elements and β-turns for human pGSHPx

A. *Secondary structure elements*[a]

Element	Residues
$3'_{10}$	12–14
$3''_{10}$	29–31
β1	35–39
α1	49–62
$3'''_{10}$	62–64
β2	65–71
α2	86–94
β3	105–107
α3	118–126
β4	157–161
β5	165–170
α4	175–184

B. β-*Turns*[b]

Type	Sequence	Residues
IV	TIDG	19–22
II	YAGK	31–34
IV	NQFG	74–77
IV	KQEP	78–81
I	ENSE	83–86
I′	GGGF	97–100
IV	GEKE	114–117
IV	GTSD	136–139
I	KVHD	147–150
IV	WNFE	153–156
I	GPDG	161–164
I	HHRT	171–174

[a] The secondary structure elements were assigned with the program DSSP (Kabsch & Sander, 1983).
[b] The β-turns were classified according to the conventions of Venkatachalam (1968) and Richardson (1981) using the program PROMOTIF (Hutchinson & Thornton, 1996).

them, α1 (residues 49 to 62), α2 (residues 86 to 94) and α4 (residues 175 to 184) are located on one side of the sheet and another one, α3 (residues 118 to 126), is on the other side. Helices α1 and α4 are nearly parallel and their axes are roughly parallel with the central β-strands. The other two α-helices, α2 and α3, are oriented perpendicular to each other and to the helices α1 and α4. Helix α1 is part of the crossover connection between the strands β1 and β2, while helix α2 is that between strands β2 and β3.

The N-terminal residues (residues 1 to 11) adopt an extended conformation, located just beside the central β-sheet and nearly parallel with the strand β5. Connected to this region is a 3_{10}-helix ($3'_{10}$, residues 12 to 14). Residues 15 to 28 form a β-hairpin that consists of two short antiparallel β-strands (residues 15 to 17 and 26 to 28) and a β-turn (residues 19 to 22) in the middle. Since both comprise two rather short strands, they are not listed in Table 1A. The β-hairpin is followed by the second 3_{10}-helix ($3''_{10}$, residues 29 to 31) and a β-turn (residues 31 to 34). Residues 35 to 39 form one of the central β-strands, β1. This β-strand together with α1 and β2 constitute a βαβ motif. The selenocysteine residue 45 is located in the loop (residues 40 to 48) connecting β1 and α1. Helix α1 and strand β2 are connected by another 3_{10}-helix ($3'''_{10}$, residues 62 to 64). The residues between β2 and α2 form an irregular segment containing three consecutive β-turns (residues 74 to 77, 78 to 81 and 83 to 86). β2, α2 and β3 constitute the second βαβ motif. Helix α2 ends with *cis*-Pro96 and is connected to β3 by a β-turn (residues 97 to 100). By another β-turn (residues 114 to 117), the chain moves to the other side of the central β-sheet and is continued by helix α3. Helix α3 is followed by a long loop connected to the two antiparallel β strands, β4 and β5. Again, three β-turns exist in this long loop, which are formed by residues 136 to 139, 147 to 150 and 153 to 156. The β-strands β4 and β5 are joined by a β-turn (residues 161 to 164) and form the second β-hairpin motif. The chain ends with the last β-turn (residues 171 to ‘174) and the C-terminal helix α4.

Of the 187 residues in a subunit, 73% were found in well-defined secondary structure elements: 19% form α helices, 14% form β-strands, 5% form 3_{10} helices and 35% form different types of β-turns. The five β-strands, which form the central core of the subunit, are predominately composed of hydrophobic residues. The central β-sheet (26 residues) is composed of only two (7%, one Arg and one Lys) charged residues, while the other residues are hydrophobic (69%, five Leu, four Val, four Phe, three Ile, one Pro and one Met), Gly (11%, three Gly) and polar (11%, one Trp, one Gln and one Tyr).

Dimer and tetramer

The residues involved in the formation of the interface between the two subunits in the dimer within the asymmetric unit (Ren *et al.*, 1995) come mainly from helix α2, the loops between β1 and α1, β2 and α2, α3 and β4. The inter-subunit contacts include several hydrogen bonds (Tyr44A OH...Tyr44B OH, Glu86A(B) OE1...Arg95B(A) NH1, Glu86A(B) OE2...Arg95B(A) NH1 and Thr90A(B) OG1...Tyr44B(A) OH) and several hydrophobic interactions, which involve residues Pro81A(B), Gly82A(B), Pro89A(B), Val94A(B) and Phe142A(B) and the hydrophobic parts of the side-chains of Tyr44A(B), Lys78A(B), Thr90A(B) and Tyr93A(B).

Two asymmetric units related by a crystallographic 2-fold axis form a tetramer in the crystal (Figure 4). The tetramer shows a flat shape with approximate dimensions of 77 Å × 75 Å × 43 Å. At the center of the tetramer there is an oblate hole with a diameter of 12 to 15 Å. Some of the residues in the long loop between α3 and β4 form the two interfaces between the dimers. These residues are located in two regions of the loop, residues 132 to 139 and 144 to 148, where the polypeptide chains run almost antiparallel. The interactions at each interface are mediated mainly by hydrogen bonds or salt bridges. Because of the 2-fold symmetry, residues between 132 and 139 in one asymmetric unit interact with residues between 148 and 144 in the other asymmetric unit and *vice versa*. Hydrogen

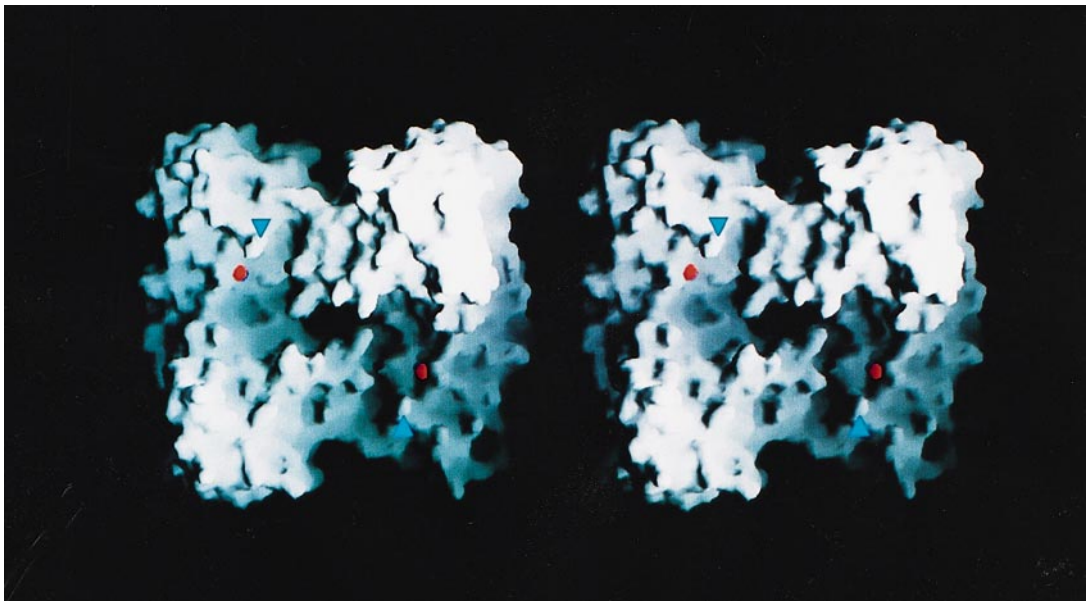

Figure 4. Molecular surface of the human pGSHPx tetramer, showing the active site pockets. The oxygen atoms bound to the selenium atom in Cso45 are shown as red spheres. The protrusion right above Cso45, indicated by a blue triangle, corresponds to Arg173. The diagram was generated with the program GRASP (Nicholls *et al.*, 1991).

bonds or salt bridges can be formed between the pairs: Glu133 O...Val′148 N, Gly136 O...Met′146 O, Asp139 OD2...Asp′139 OD2, Asp139 OD1...Glu′144 OE2, Glu144 OE1...Ser′138 OG. Other residues found at the interface are Ser132, Leu134, Leu135, Thr137, *cis*-Pro145 and Lys147.

The active site selenocysteine residues are found in tetradic arrangement. The distance between the two selenium atoms, related by the local dimer axis, is 23.2 Å. The Se-Se distance of the selenocysteine residues, related by the crystallographic 2-fold axis, is 37.6 Å. The selenocysteine residues are located on two opposite surfaces of the tetramer. Two selenocysteine residues on either surface are related by an additional 2-fold axis, generated by the local 2-fold and crystallographic 2-fold. The distance between those selenium atoms is 41.1Å.

The solvent-accessible surface areas, interface areas and volumes are given in Table 2.

Active site

The active site residues in the environment of the selenocysteine residue (Cso45) are generally well defined in both ($2F_o - F_c$) and averaged electron density maps (see Figure 2). Inspection of difference ($F_o - F_c$) electron density maps calculated with phases from the refined model indicated that the selenium atom in Cso45 exists in an oxidized state, as a seleninic acid (R-Se(O)OH), with two oxygen atoms bonded to the selenium atom (see Figure 10 and Discussion). Upon dimer formation in the asymmetric unit, a pocket is formed on the protein surface, as shown in Figure 4. Cso45 and its neighboring residues are located in this surface pocket. As Cso45 is situated near the interface of the two subunits, the pocket is formed by residues from both subunits. Most of these residues are located in various loops. Residues 132 to 140 form a part of the pocket, which is also a part of the interface between asymmetric units in the tetramer. The N-terminal residue Lys1 and residues 171 to 173, 46 to 47 and 95 form another part of the pocket. Cso45 is located at the bottom of the pocket opposite the part formed by residues 132 to 140. The protrusion right above Cso45 on the protein surface corresponds to Arg173. The other half of the surface pocket is formed by residues from the other subunit. These residues are 83′ to 86′ in a β-turn, 108′, 18′, 24′ to 25′ and 116′ to 117′ in loops and 118′ to 120′ in helix α3.

The bottom of the pocket is formed mainly by residues in loops connecting α-helices and β-strands. A large part of them are residues in the long loop between α3 and β4, while the others are in the loops between β3 and α3, β2 and α2, β5 and

Table 2. The solvent-accessible surface areas, interface areas and volumes for human pGSHPx

	Accessible surface area (Å^2)	Interface area (Å^2)	Volume (Å^3)
Monomer	9222	/	19,240
Dimer	17,391	1052	38,470
Tetramer	32,025	4862	76,880

The solvent-accessible surface areas were calculated with a probe radius of 1.4 Å.

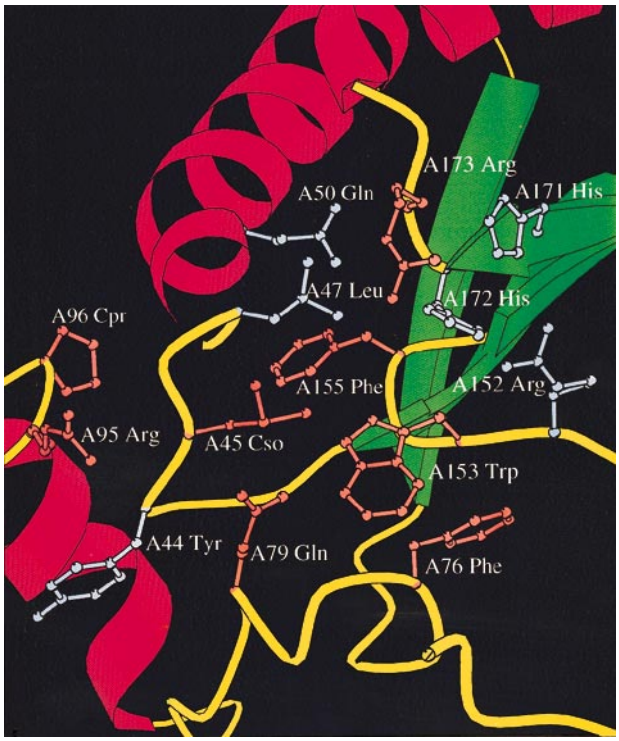

Figure 5. Residues in the neighborhood of seleocysteine residue Cso45, which is located in a loop turn between β1 and α1. In comparison to the bovine erythrocyte enzyme, the conserved residues are coloured in magenta and the mutated residues are coloured in blue. The Figure was generated with MOLSCRIPT (Kraulis, 1991).

α4, β1 and α1 and several residues in helix α3. The bottom around Cso45 is formed mainly by hydrophobic residues (Phe76, Trp153, Phe155, Tyr44, Leu47, Trp143, Phe142 and Leu141), while the remaining residues forming the bottom surface are more negatively charged (see Figure 9). The active site pocket is open to bulk solvent. The size of the pocket is about 15 Å (Arg95 NH2 – Arg140 NH1) × 44 Å (Lys1 NZ – Tyr25′ OH) with a varying depth of 6 to 8 Å around the center of the bottom.

Cso45 is located in a turn connecting β1 and α1. The active site residues neighboring Cso45 are shown in Figure 5. Close to Cso45, there are several aromatic residues, including Tyr44, Phe76, Trp153, Phe155, His171 and His172. Cso45 forms a hydrogen bond network with its neighboring residues. A hydrogen bond is formed between Cso45 and Trp153 (Cso45 OD1...Trp153 NE1, 2.9 Å). Oxygen atom OD2 of Cso45 can form hydrogen bond interactions with two main-chain nitrogen atoms, Cso45 OD2...Gly46 N, 2.5 Å and Cso45 OD2...Leu47 N, 3.1 Å. This oxygen atom (OD2) is more exposed than OD1 to the protein surface. Another residue, Asn154, not visible in Figure 5, is within hydrogen bonding distances of both Cso45 and Trp153 (Cso45 OD1...Asn154 OD1, 3.1 Å; Trp153 NE1...Asn154 OD1, 2.8 Å). As Arg95 is close to Cso45, there may be hydrogen bonding interactions too, Cso45 N...Arg95 NH1, 3.8 Å. The selenium atom in Cso45 is furthermore within hydrogen bonding distances to Gln79 and Trp153 (Cso45 SE...Gln79 NE2, 3.5 Å; Cso45 SE...Trp153 NE1, 3.6 Å). These two residues are conserved in the bovine erythrocyte enzyme and it has been suggested that these hydrogen bonds may help to stabilize the selenolate anion ($R\text{-}Se^{\ominus}$) during catalysis, although the actual mechanism catalyzed by the cellular enzyme is still not well understood (Epp *et al.*, 1983). Other residues near Cso45 include Leu47, Gly49, Gln50, Arg152 and Arg173. As shown in Figure 4, Arg173 is located right above the selenocysteine side-chain on the protein surface. Table 3 lists the possible hydrogen bonding interactions between residues at the active site.

Table 3. Hydrogen bonding interactions in the active site

Atom 1	Atom 2	*d* (Å)
Tyr44 O	Arg95 NH1	2.9
Cso45 N	Arg95 NH1	3.8
Cso45 SE	Gln79 NE2	3.5
Cso45 SE	Trp153 NE1	3.6
Cso45 OD1	Trp153 NE1	2.9
Cso45 OD1	Asn154 OD1	3.1
Cso45 OD2	Gly46 N	2.5
Cso45 OD2	Leu47 N	3.1
Leu47 O	Gln50 NE2	3.4
Gln50 NE2	His172 O	3.4
Gln79 NE2	Trp153 NE1	3.5
Gln79 NE2	Asn154 OD1	2.6
Trp153 NE1	Asn154 OD1	2.8
His171 ND1	His172 N	2.8
His171 ND1	His172 ND1	3.4
His171 NE2	Arg173 N	3.3

Superposition of the two subunits in the asymmetric unit indicates that most of the residues within the active site show similar conformations except the two arginine residues Arg173 and Arg152. Their side-chains adopt different conformations in the two subunits, possibly responding to different packing interactions in the crystal (Figure 6). For both arginine residues, clear and well-defined electron densities were observed in both subunits. In subunit A, the side-chains of the two residues are oriented towards each other and their guanido groups are within distances for non-covalent interactions. In subunit B, inter-tetramer interactions between this subunit and its symmetry-related counterpart (B′) have been detected as a result of crystal packing. B′ is positioned near the active site pocket of subunit B. The interacting residues are located mainly around the active site of subunit B and in two additional loops (loop in the first β-hairpin, loop between α2 and β3) of B′. These residues are Gly46, Leu47, Glu53, His172, Arg173, Tyr175, Leu18′, Glu23′, Glu24′, Tyr25′ and Glu108′. Several hydrogen bonds or salt bridges could be formed between His172, Arg173 and Glu24′, Glu108′: His172 O...Glu24′ OE1, 3.2 Å; Arg173 NH1...Glu24′ OE1, 3.4 Å; Arg173 NH2...Glu24′ OE2, 3.9 Å; Arg173 NH2...Glu108′

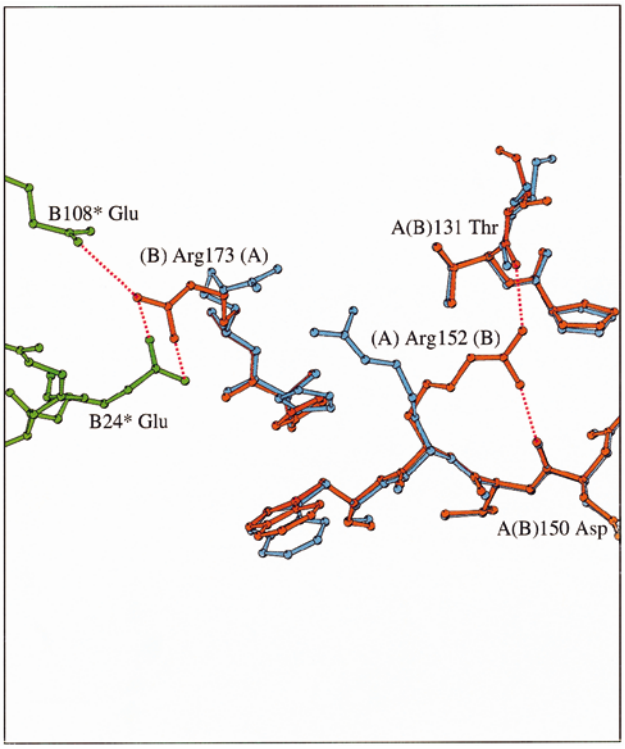

Figure 6. Superposition of subunits A and B showing the alternative side-chain conformations of the two arginine residues, Arg152 and Arg173. In subunit A (blue), the side-chains of the two arginine residues were oriented towards each other. In subunit B (red), their side-chains were oriented away from each other in order to form hydrogen bonds or salt bridges with their neighboring residues. The symmetry-related molecule near subunit B is colored green.

OE2, 3.3 Å; Arg173 O...Glu24′ OE2, 3.2 Å. To form hydrogen bonds (salt bridges) with Glu24′ and Glu108′, the side-chain of Arg173 is moved away from Arg152 and is oriented towards the carboxylate group of Glu24′. Part of the side-chain of Arg152, including the guanido group, appears rotated by almost 180° and can thus form two hydrogen bonds with the amino oxygen atoms of Glu133 and Asp150: Arg152 NH1...Thr131 O, 3.0 Å; Arg152 NH2...Asp150 O, 3.1 Å. As these two residues are located near the ends of the long loop between α3 and β4, the formation of both hydrogen bonds could contribute to the conformational stability of the loop.

As mentioned before, a segment (residues 132 to 140) of the loop between α3 and β4 forms part of the active site pocket, which is located on the protein surface just opposite Cso45. The residues in this segment are generally well defined in the electron density map, except those from Gly136 to Asp139. The averaged electron density map showed diffuse densities for the polypeptide chain in this region, suggesting alternative conformations. Therefore the main chain may flip back and forth with respect to Cso45. This movement could change the size of the active site pocket. The maximum distance between C^{α} atoms in the alternative conformations is about 6.5 Å. The flexibility of this segment might be partially due to the existence of Gly136.

Discussion

Comparison with bovine erythrocyte glutathione peroxidase

Glutathione peroxidase is member of a large enzyme/protein family (Ursini *et al.*, 1995), but only two crystal structures have been determined of the enzymes in this family. In addition to the human plasma GSHPx, the other known structure is that for a cellular enzyme, bovine erythrocyte GSHPx (Epp *et al.*, 1983). Both structures have been solved in their unliganded forms. Although they belong to different species, their subunit structures are quite similar (Figure 7) as expected from their sequence identity (42%). No significant deviation of secondary structure elements was observed between the two enzymes. Superposition of the two structures exhibits an r.m.s. deviation of 1.43 Å for 178 C^{α} atoms. A structure-based sequence alignment of the two enzymes is shown in Figure 8.

One of the major differences between the two structures is at the N terminus. Compared with the

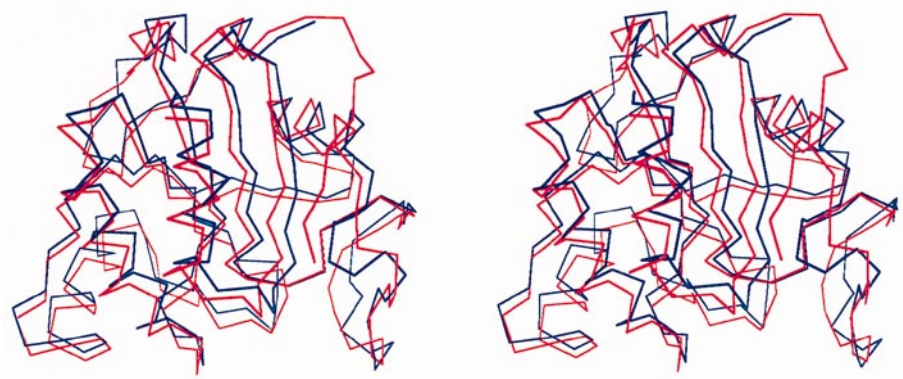

Figure 7. Stereo view of the C^{α} superposition of human plasma GSHPx (red) and bovine cellular GSHPx (blue). Stereo view. The Figure was produced with MOLSCRIPT (Kraulis, 1991).

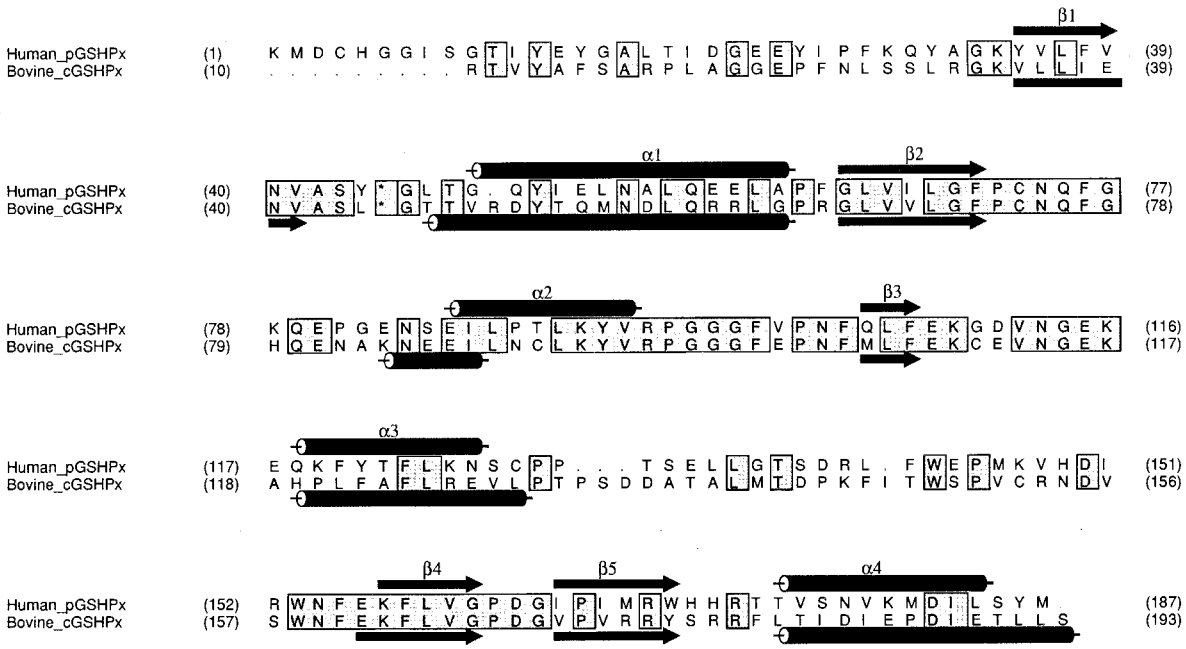

Figure 8. Structure-based sequence alignment of human plasma (human_pGSHPx) and bovine cellular (bovine_cGSHPx) glutathione peroxidases. α-Helices are indicated by cylinders, β-strands by arrows, conserved residues by shaded boxes. The selenocysteine residues (Cso45) are represented by asterisks (*). The residue numbers are indicated in parentheses. The Figure was prepared with ALSCRIPT (Barton, 1993).

bovine cellular enzyme, the human plasma enzyme has nine extra N-terminal residues. These residues form a rather extended segment, which is parallel with β5 like an additional β-strand in the central β-sheet. This segment is packed tightly to the protein surface between helix α4 and a part of the long loop connecting α3 and β4, while in the bovine cellular enzyme this space is occupied mainly by a protruding portion of the long loop. The N-terminal residue, Lys1 corresponding to the 29th residue in the amino acid sequence derived from cDNA (Takahashi *et al.*, 1990), forms part of the active site pocket. Its side-chain is within hydrogen bonding distance to Glu133 in the loop. All the N-terminal residues are well defined in the electron density map. There was no possibility to extend the chain from Lys1 because of tight crystal packing. Whether this lysine residue represents the actual N terminus of he native protein is still an open question. A predicted signal peptide cleavage site (Von Heijne, 1986) is likely between the 24th and 25th residue in the primary structure derived from cDNA (Takahashi *et al.*, 1990). The cleavage of this signal peptide would result in a protein with four more N-terminal residues. Unfortunately, attempts to determine the N terminus of the native protein have been unsuccessful so far (Avissar *et al.*, 1991; Esworthy *et al.*, 1991). Ambiguity also exists for the amino terminus of the bovine cellular enzyme. Compared to the amino acid sequence derived from cDNA (Mullenbach *et al.*, 1988), 16 N-terminal residues were missing in the crystal structure of bovine erythrocyte GSHPx. A primary structure determination by Günzler *et al.* (1984) suggested that nine N-terminal residues were truncated by proteolysis in the protein purified for crystallographic studies and the other seven residues formed a segment that was assumed to be processed to give the mature protein. However, a later study (Gettins *et al.*, 1992) indicated that the normal form of the bovine cellular enzyme is most likely the full transcript with the terminal methionine residue removed and the adjacent cysteine residue *N*-acetylated. The proteolysis by contaminating proteases during purification resulted in the formation of different forms of the enzyme, including that characterized by Günzler *et al.* (1984) and that used in crystal structure determination. A similar situation was observed in the purification of ovine erythrocyte GSHPx. It should be noted that activity measurements on the intact and truncated forms showed no significant differences. Thus it has been suggested that the 16 residues missing in the crystal structure of the bovine cellular enzyme may form an N-terminal tail that is susceptible to proteolysis and is attached to the globular core of the protein. As this is an alanine-rich region (9 of 16), it has been predicted to be largely α-helical (Chou & Fasman, 1978). In the crystal structure of the human plasma enzyme, however, the corresponding N-terminal residues adopt a β-strand-like conformation.

Some C-terminal residues are absent from both crystal structures in comparison with the sequences derived from cDNA. Those residues (11 in the human plasma enzyme and five in the bovine cellular enzyme) are located after the corresponding helices α4 in both structures. The missing residues

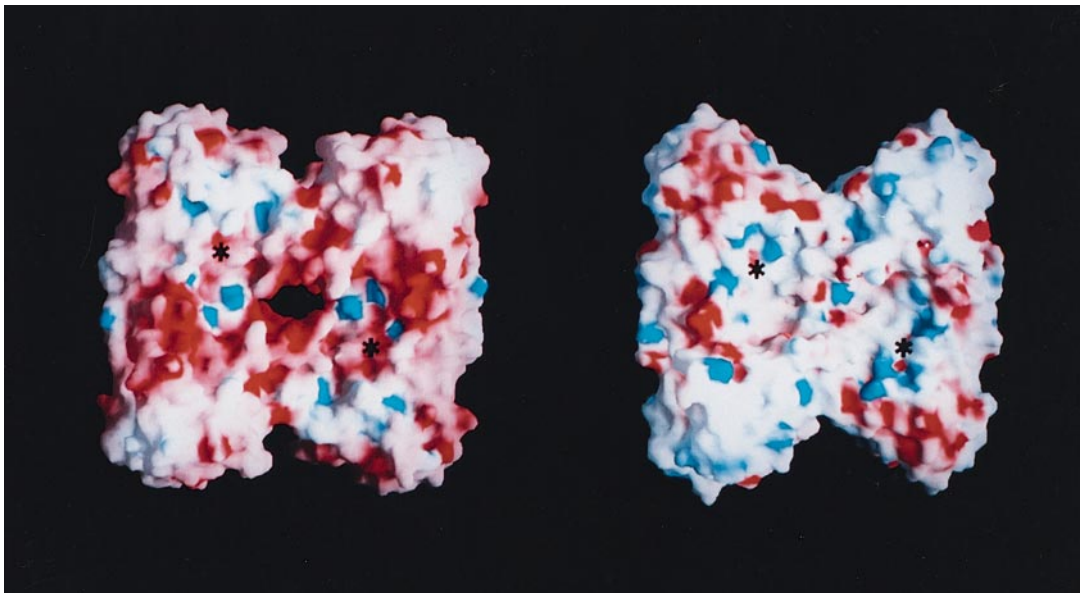

Figure 9. Electrostatic surface potential of human plasma GSHPx (left) and bovine cellular GSHPx (right) tetramers. Both maps are contoured within the same potential range from −12.5 to 7.3 (arbitrary units) with positively charged regions in blue and negatively charged regions in red. The positions of selenocysteine residues are indicated by asterisks (*). The Figure was produced with GRASP (Nicholls *et al.*, 1991).

in the bovine cellular enzyme were obviously not observed during model building in which both averaged and $(2F_o - F_c)$ maps were used (Ladenstein *et al.*, 1979; Epp *et al.*, 1983). Although the averaged electron density map did not show clear densities for the missing residues in the human plasma enzyme, short pieces of densities after helices α4 were indeed observed within both subunits in the $(2F_o - F_c)$ electron density map. Since these electron densities suggested different chain directions, no attempt was made to include them in the current model because of the limited resolution of the diffraction data. Whether the poor quality of the averaged map resulted from the flexibility of the C termini or whether some of the residues were cleaved during purification needs further investigation.

Another major structural difference between the two enzymes was observed in the longest loop: the loop between α3 and β4. In the human plasma enzyme, the residues in this loop form the interfaces between the dimers of the tetramer and a part of the active site pocket that is located on the protein surface opposite the selenocysteine residue. A structural comparison between the two enzymes indicated that four of the five deletions in the human plasma enzyme occured in this loop near the residues that form a part of the active site pocket. As a result, a protrusion of the loop (residues 131 to 136) existing in the bovine cellular enzyme disappeared and the spare space is occupied by the extented N-terminal residues. In the human plasma enzyme, the shortened loop was generally well defined in the electron density maps, except the segment between Gly130 and Asp139 (see Results). The conformation of this loop is stabilized by three successive β-turns (Table 1B) and hydrogen bonding interactions, such as those at the interfaces between the dimers. An unusual occurrence of two consecutive proline residues (Pro129-Pro130) in the human plasma enzyme could reduce the conformational freedom of the loop. The shortened loop may be further stabilized by the formation of a disulfide bridge between Cys4 and Cys128.

While not observed in the bovine cellular enzyme, a disulfide bridge probably exists in the human plasma enzyme. The primary structure of the plasma enzyme derived from cDNA contains four cysteine residues. One of them is most likely within the signal peptide (ninth in the sequence) and thus not present in the model. In the subunit structure, two of the remaining three cysteine residues, Cys4 and Cys128, which are both not conserved in the cellular enzyme, are spatially very close to each other and may form a disulfide bridge. The $(2F_o - F_c)$ electron density map showed a density connection between the two sulfur atoms (Cys4 SG...Cys128 SG, 4.3 Å) in subunit B. However, they deviated away slightly in subunit A (Cys4 SG...Cys128 SG, 6.6 Å). Those disulfide bridges could form quite easily in both subunits with little adjustment. The storage of the plasma enzyme in a solution containing 1 mM DTT before crystallization might lead to the observed uncertainty in the structure. The existence of intrachain disulfide bridges in the human plasma enzyme has been suggested (Avissar *et al.*, 1989b). The oxidized plasma enzyme showed higher mobility in SDS-PAGE than its reduced form, while this was not observed for the cellular enzyme. The formation of the disulfide bridge is

likely to enhance the conformational stability of the plasma enzyme, especially at the N-terminal region and the long loop.

Unlike the cellular enzyme, it is generally accepted that the extracellular plasma enzyme is a secreted glycoprotein (Takahashi *et al.*, 1987; Avissar *et al.*, 1994). It has been shown that the plasma enzyme, but not the cellular enzyme, can be stained by [^{125}I]concanavalin A. The observation, that most of the stain (90%) was removed by peptide *N*-glycohydrolase F, has suggested, that the protein is most probably *N*-glycosylated. However, other kinds of linkages, such as *O*-glycosylation, have not been ruled out completely (Avissar *et al.*, 1989a). Inspection of the amino acid sequence of the plasma enzyme, however, did not reveal the sequence motif NXT(S), where *N*-glycosylation at the ND atom of Asn usually occurs. Excess density, possibly corresponding to carbohydrate, was not observed in the electron density maps. As a consequence, no sugar linkage was assigned in the current model and the exact type of glycosylation, if any, is not yet established.

The structural differences between the two enzymes increase with increasing quaternary structure. The r.m.s. deviations upon superpositions of the dimers and tetramers are 1.59 Å (356 C^{α} atoms) and 2.06 Å (712 C^{α} atoms), respectively. Sequence alignment indicated that the residues that form the subunit interfaces in both enzymes are located in the same topological regions. However, a reduction of hydrogen bonds between the dimers in the tetramer of the plasma enzyme was observed. This is mainly due to the shortening of the long loop between α3 and β4, which forms the dimer interfaces.

The tetramer of the bovine cellular enzyme has a volume of 75,700 Å^3 with a solvent-accessible area of 28,956 Å^2. These values are 1.5% and 9.6% smaller than those of the plasma enzyme, respectively. However, a 31.3% larger buried area upon subunit aggregation, 6383 Å^2, was observed for the cellular enzyme. The observation that less residues were buried in the plasma enzyme subunit interfaces is largely due to the shortening of the loop between α3 and β4. Furthermore, a hole was observed at the center of the tetramer of the plasma enzyme (Figure 9). In the cellular enzyme, this place is occupied mainly by the side-chains of Lys144 contributed by the four subunits.

The subunit structures of both plasma and cellular enzymes showed the characteristic structural motif, which consists of a central β-sheet and flanking α-helices. Similar structures have been found in other redox proteins, including thioredoxins, glutaredoxins, glutathione S-transferase and DsbA (the protein catalyzing disulfide formation *in vivo*) (Martin, 1995). All these proteins interact with substrates having a thiol or a disulfide group. Based on the alignment of their structures, a common motif, namely the thioredoxin fold, has been defined (Martin, 1995). It consists of a four-stranded β-sheet and three flanking α-helices. The additional structural elements in each of the proteins are located at certain insertions within the thioredoxin fold. For glutathione peroxidases, two large insertions are observed. One is located between β2 and α3, including α2 and β3, the other one is the long loop between α3 and β4. The subunit interfaces in the tetramers of both enzymes are constituted almost entirely by residues in these two insertions. Compared to the bovine cellular enzyme, the second insertion constitutes the least conserved region of the plasma enzyme, where almost all the observed deletions occured. The structural similarity among these redox proteins is striking considering the fact that there is no obvious sequence similarity among them. On alignment of their structures, it has been found that the important active site residue in each of the proteins, which is the selenocysteine residue in glutathione peroxidase, is located at the same relative position in space. Thus, these proteins might have evolved from a common ancestor with less specificity in its substrate binding interactions (Sinning *et al.*, 1993; Martin, 1995). However, convergent evolution cannot be ruled out.

In addition to the human plasma enzyme, other plasma glutathione peroxidases have been characterized in mammals, which include the rat plasma enzyme (Yoshimura *et al.*, 1991), bovine plasma enzyme (Martin-Alonso *et al.*, 1993) and mouse plasma enzymes (Master *et al.*, 1994; Schwaab *et al.*, 1995). The amino acid sequences of these plasma enzymes derived from cDNA share a pairwise identity of more than 87%. For each pair, more than one-third of the mutated residues are located near both termini in the regions that were not observed in the electron density map of the human plasma enzyme. The pairwise identity is around 92% if these residues are not included. Almost all of the residues at the active site close to the selenocysteine residue are conserved in these enzymes. Based on the high level of identity of their amino acid sequences, the overall fold and active site architecture of the other plasma enzymes can be expected to show a high level of similarity to those of the human plasma enzyme.

Active site

An unusual characteristic of glutathione peroxidases is the existence of a selenocysteine residue in its active site. Biochemical (Flohé, 1989), kinetic (Flohé *et al.*, 1972) and crystallographic (Epp *et al.*, 1983) studies on the bovine cellular enzyme have suggested that the reduction of hydroperoxide by GSHPx is a cyclic process and that the selenocysteine residue is involved in the catalytic cycle. It has been suggested that the selenium atom undergoes two valence states in the cycle, the selenolate anion (R-Se$^{\ominus}$) and the selenenic acid (R-Se-OH) states. Upon reduction of hydroperoxides, the selenolate anion (R-Se$^{\ominus}$) is oxidized to selenenic acid (R-Se-OH), which is subsequently reduced again by GSH. However, under high concentrations of oxidizing substrates, the selenenic acid (R-Se-OH) can be oxidized to a third state, the seleninic acid

(R-Se(O)OH), as observed in the crystal structure. This reaction presumably does not happen under physiological conditions, however, upon air oxidation it may lead to a ''resting form'' of the enzyme. Selenium in the selenocysteine residue acts as a strong nucleophile in catalytic reactions. It is well known that organic selenium compounds are more reactive as nucleophiles than their sulfur counterparts (Odom, 1983). The lower redox potential of selenocysteine compared with cysteine is catalytically favorable for the reaction catalyzed by glutathione peroxidase. At neutral pH, selenols (in contrast to thiols) exist mostly in anionic form and represent good reducing groups under normal physiological conditions (Mills, 1957; Stadtman, 1987).

As a strong X-ray scatterer, the selenium atom in the selenocysteine residue showed the highest electron density in both $(2F_o - F_c)$ and averaged maps of the human plasma enzyme. The valence state of the selenium atom in the plasma enzyme was investigated by inspecting a $(F_o - F_c)$ omit map in addition to $(2F_o - F_c)$ and averaged electron density maps. The selenium atom was found to exist as a seleninic acid (R-Se(O)OH) in the crystal. The $(F_o - F_c)$ omit electron density map (Figure 10) shows clearly density features that can be interpreted as two oxygen atoms bound to the selenium atom. The same oxidation state of selenium in the selenocysteine residue has been reported in the study of the bovine cellular enzyme (Epp *et al.*, 1983), however, in this case after treatment with 5 mM H_2O_2.

Easy changes of the oxidization state is a characteristic property of selenium. Studies of organic selenium compounds (Odom, 1983) indicated that selenium in comparison to sulfur can be more easily oxidized from valence state (II) to (IV) but changes from (IV) to (VI) are more difficult to achieve. As a result, selenides (R-Se-R′) and selenenic acids (R-Se-OH) are converted rather to selenoxides (R-SeO-R′) and seleninic acids (R-Se(O)OH), than to selenones (R-SeO_2-R′) and selenonic acids (R-Se(O_2)OH. In general, compounds with terminal selenium atoms easily degrade upon air treatment, whereas those with selenium in valence states (II) and (IV) are relatively stable. The seleninic acid state (R-Se(O)OH) that we have detected in the human plasma enzyme is possibly due to air oxidation in the experimental process, for example the long storage of the protein during purification and crystallization. It does certainly not represent an artificially overoxidized form of the enzyme, but rather a resting form that may exist in an oxidizing environment. High susceptibility of selenium to oxidation was observed for selenomethionine residues in a crystallographic study of recombinant *Escherichia coli* selenomethionyl thioredoxin (Hendrickson *et al.*, 1990).

The overall active site architecture of the human plasma enzyme is similar to that of the bovine cellular enzyme. No remarkable deviation of the peptide backbone was observed in the active site region. However, details of the environment close

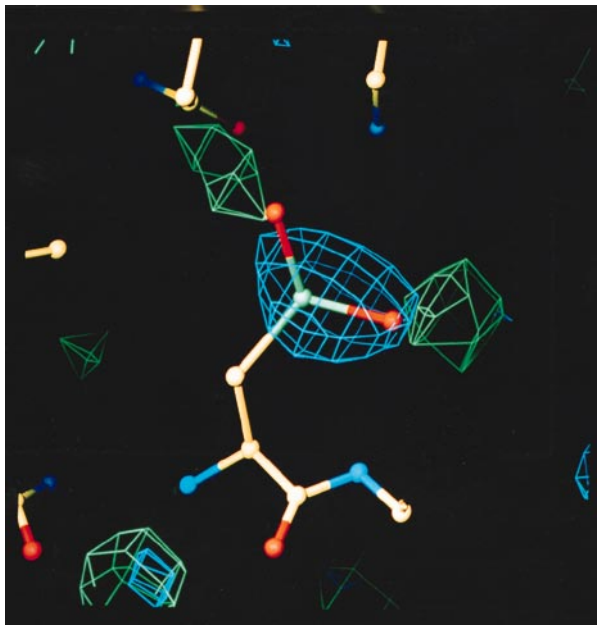

Figure 10. $(F_o - F_c)$ difference electron density maps showing the selenium atom and indicating its oxidization state in the native crystals. The map colored in blue was calculated with structure factors from the final model but with the selenium atom excluded from the calculation. In the map coloured in green, the two oxygen atoms bonded to the selenium were excluded from structure factor calculations. Both maps are contoured at the 2σ level.

to the selenocysteine residue are quite different between the two enzymes. Approximately only half of the residues close to the selenocysteine residue within a range of 10 Å are conserved in both enzymes. The residues conserved in the human plasma enzyme are Phe76, Gln79, Arg95, Trp153, Phe155, Asn154 and Arg173. Of these residues, Gln79 and Trp153 are located within hydrogen bonding distance to the selenium atom and have been suggested to play functional roles in catalysis (Epp *et al.*, 1983). A further comparison of amino acid sequences indicated that these two residues are in fact conserved in the whole glutathione peroxidase superfamily, which consists of 21 glutathione peroxidases and their homologs (Ursini *et al.*, 1995). This implies that they share similarities in their catalytic mechanisms. Compared with the bovine cellular enzyme, the mutated residues in the human plasma enzyme close to Cso45 are Tyr44, Leu47, Gly49, Gln50, His171, His172 and Arg152. A cluster of aromatic residues on one side of Cso45 was observed in the human plasma enzyme, which includes Tyr44, Phe76, Trp153, His171 and His172 (Figure 5). These residues form an aromatic network. The distance between the centroids of neighboring aromatic rings ranges from 4.2 to 5.9 Å. These aromatic residues are conserved in the other plasma enzymes found to date (Yoshimura *et al.*, 1991; Martin-Alonso *et al.*, 1993; Master *et al.*, 1994;

Schwaab *et al.*, 1995) except a His171 to Tyr mutation.

Based on the crystal structure of the bovine cellular enzyme, a hypothetical model of glutathione binding was proposed by Epp *et al.* (1983). In this model, three residues are involved in the binding of glutathione. Arg177 and Arg50 form salt bridges with the γ-glutamyl carboxyl group and C-terminal glycyl carboxyl group of glutathione, respectively, whereas the residue at position 140 forms a hydrogen bond with the N-terminal amino group. Residue 140 was originally assigned to a Gln in a tentative amino acid sequence, but was shown o be Met afterwards (Günzler *et al.*, 1984). A structural comparison between the bovine cellular enzyme and human plasma enzyme indicated that all these three residues, however, are mutated or deleted in the plasma enzyme. Arg177 and Met140 are mutated to His172 and Gly136, while Arg50 is deleted in the plasma enzyme.

Inspection of these changes revealed interesting structural details. The deleted residue Arg50 does not have a structural counterpart in the plasma enzyme. The only possible alternative is Gln50 (plasma) mutated from Asp51 (cellular), whose position shifted closer to Cso45 because of a shortening of the polypeptide chain. A mutation of Met140 (cellular) to Gly136 (plasma) was observed in the long loop between α3 and β4. Badly defined electron densities were observed for this glycine residue and its neighboring residues, including Thr137, Ser138 and Asp139, which may suggest flexibility of this fragment. Since it forms part of the active site pocket and is located opposite to Cso45, it seems interesting to further investigate whether this fragment plays a role in substrate binding.

Arg173, which is located next to the mutated residue His172 (plasma), has a flexible side-chain that adopts alternative conformations in the two subunits within a dimer of the asymmetric unit (Figure 6). A concerted conformational change for the side-chain of Arg152 was observed in relation to Arg173. More interestingly, it seemed that the conformational change of Arg173 was induced by a glutamic acid residue in a crystallographically symmetric neighboring molecule of subunit B. It is tempting to speculate that a similar situation may occur when the plasma enzyme binds the reducing substrate, such as GSH, which contains a γ-glutamyl carboxyl group. This observation, to some extent, coincides with the binding model proposed for the cellular enzyme. In that model a salt bridge is suggested between the γ-glutamyl carboxyl group of GSH and Arg177 (cellular), which is mutated to His172 in the plasma enzyme.

Unlike the cellular enzyme, the biological function of the human plasma enzyme remains still unclear. Enzymatic studies using GSH as the reducing substrate revealed it as an enzyme with low activity (Takahashi *et al.*, 1987; Maddipati & Marnett, 1987; Yamamoto & Takahashi, 1993; Esworthy *et al.*, 1993). Although the plasma enzyme can reduce hydrogen peroxide and organic hydroperoxides, it is approximately tenfold slower than the cellular enzyme. It may be able to react with certain phospholipid hydroperoxides, but the reactions were usually either incomplete or less effective than those catalyzed by the phospholipid enzyme. While the cellular enzyme does not show saturation with respect to GSH (Flohé *et al.*, 1972; Forstrom *et al.*, 1978a), the plasma enzyme exhibits saturation kinetics with a K_m value of 4.3 mM to 5.3 mM (Takahashi *et al.*, 1987; Maddipati & Marnett, 1987). Since the concentration of reduced thiol groups in human plasma is only 0.3 to 4 μM, it is quite unlikely that GSH is the reducing substrate for the human plasma enzyme. Even though all the peroxidase activity in human plasma can be attributed to the plasma enzyme (Maddipati *et al.*, 1987), it is still difficult to assign it a definite physiological role. However, the low reactivity of the plasma enzyme with GSH and the low levels of reduced thiol groups in plasma may suggest that other reducing substrates are used by this enzyme. The experiment by Björnstedt *et al.* (1994) showed that extracellular thioredoxin reductase, thioredoxin or glutaredoxin in human plasma could be reasonable candidates. In this experiment, the plasma enzyme has shown a higher affinity (more than three times) for the thioredoxin system than the cellular enzyme.

The observed structural differences between the human plasma and bovine cellular enzyme, especially at the active site, may account for possible differences in the binding of substrates. The conformational flexibility of certain active site residues described above was not observed in the study of the bovine cellular enzyme, which, by contrast, suggested that the conformations of the active site residues are rather rigid (Ladenstein *et al.*, 1979; Epp *et al.*, 1983). Furthermore, a remarkable difference was observed between the two enzymes in their electrostatic potentials on the tetramer surfaces. As shown in Figure 9, the surface of the plasma enzyme is more negatively charged, especially around the center of the tetramer and close to the active site. This difference may also affect substrate binding. However, the significance of these differences and the actual roles of the active site residues in catalysis can be elucidated only with the help of an increased structural and biochemical data basis, such as the structure of an enzyme/substrate complex and the results from mutagenesis studies, which are in progress.

Materials and Methods

Protein purification and crystallization

GSHPx was purified from human plasma (obtained from the Blood Bank, Lund University Hospital) using a modification of a previously described procedure (Huang & Åkesson, 1993). The purified protein (purification 10,000 to 20,000-fold) was concentrated to 10 to 17 mg/ml and kept in 10 mM Tris-HCl (pH 7.5), 1 mM dithiothreitol. The protein was crystallized by the sitting-drop vapor-diffusion method (Ren *et al.*, 1995). The crys-

tals diffracted to beyond 2.9 Å and belong to the tetragonal space group $I4_1$ with cell dimensions of $a = b = 82.9$ Å and $c = 130.9$ Å. Assuming one dimer per asymmetric unit, the V_m value of the crystal is 2.3 Å^3/Da, which is within the most probable range of values found for proteins (Matthews, 1968).

X-ray data collection and processing

X-ray diffraction data were collected using a MAR research image-plate system on a Siemens/Mac Science rotating-anode generator operated at 45 kV and 90 mA with CuKα radiation. The data were processed by the MARXDS program (Kabsch, 1988) and scaled and merged by the PROTEIN package (Steigemann, 1974). Care was taken to avoid inconsistent indexing choices among different datasets due to the multiple indexing possibilities of this space group. The final dataset, consisting of 8101 unique reflections with $I_o > 2\sigma(I_o)$, is 82% complete to 2.9 Å resolution and still 51% in the 3.0 to 2.9 Å shell. The merging *R*-factor for symmetry-related reflections of this dataset is 0.118. Data collection statistics are presented in Table 4.

Self-rotation function and packing model

A rotation function analysis (Rossmann & Blow, 1962) was performed to check the local symmetry of the molecules in the unit cell using the real space search options SELF of the XPLOR program suite (Brünger, 1992a). Inspection of the self-rotation function (Ren *et al.*, 1995) indicated the presence of two non-crystallographic 2-fold rotation axes that are perpendicular to each other and to the crystallographic 4-fold screw axis. These 2-fold local axes relate the monomers within plasma GSHPx dimers. Based on the self-rotation function analysis, a hypothetical model describing the packing of plasma GSHPx tetramers has been proposed (Ren *et al.*, 1995). This model, which has been verified through structure solution, not only provided the packing information for the molecules, but also facilitated the interpretation of cross-rotation and translation search results.

Patterson search calculations

The structure was solved by Patterson search methods with the XPLOR program package. The search model was a polyalanine dimer modified from the 2.0 Å crystal structure of bovine erythrocyte GSHPx (Epp *et al.*, 1983), whose amino acid sequence shows 42% identity with that of human plasma GSHPx. The cross-rotation search was carried out with 10 to 3.0 Å reflection data using inner and outer integration radii of 5 and 20 Å. The search model was placed in a cubic cell with edges of 150 Å. The correct solution was the highest peak that corresponded to the orientation of the local axis *B*. The height of this peak was 23.1 σ with the next peak at 22.4 σ. A cross-rotation search using PROTEIN gave the same solution. After Patterson correlation (PC) refinement (Brünger, 1990), this peak retained its highest position with Eulerian angles (θ_1, θ_2, θ_3) of (96.5°, 137.9°, 106.1°). The subsequent translation search was calculated for the whole unit cell using the refined orientation. The inclusion of low-resolution data from 20 to 10 Å in the translation search resulted in improved signal-to-noise ratios for detected peaks. The translation search with 20 to 3.0 Å resolution data yielded an unambiguously correct solution that is also the highest peak (10.4 σ above the mean value). The position of this peak (Figure 11), with fractional coordinates $X = 0.771$ and $Y = 0.048$, whereas Z is free, coincided very well with the center of gravity of the dimer as suggested by the packing model (Ren *et al.*, 1995). Thus, a plasma GSHPx tetramer can be generated from the dimer by applying the crystallographic symmetry operation $(-x, -y, z)$. The search model, appropriately rotated and translated in the unit cell, gave an initial *R*-factor of 43.4% for data between 8.0 and 3.0 Å.

Rigid-body refinement

The resulting rotated and translated model was subjected to rigid-body refinement with XPLOR using data from 8.0 to 3.0 Å. The two subunits in the dimer were refined with 20 cycles as two rigid groups. This was

Table 4. X-ray data collection and refinement statistics

A. *Crystal parameters*	
Space group	$I4_1$
Cell dimensions (Å)	$a = b = 82.9$, $c = 130.9$
Asymmetric unit	2 subunits
B. *Data collection*	
X-ray source	CuKα, λ = 1.542 Å
Maximum resolution (Å)	2.9
Measured reflections ($I_o > 2\sigma(I_o)$)	37,813
Unique reflections	8101
R-merge[a]	0.118
Completeness (%)	82 (8.0 − 2.9 Å); 51 (3.0 − 2.9 Å)
C. *Refinement*	
Resolution range (Å)	8.0 − 2.9
No. of non-hydrogen atoms	3004
NCS restraints (kcal mol^{-1} Å^{-2})	300 (main chain); 5 (side chain)
R-factor[b]	22.8%
Free *R*-factor	33.5%
r.m.s.d. bond length (Å)	0.017
r.m.s.d. bond angles (°)	2.1
r.m.s.d. dihedrals (°)	26.9
r.m.s.d. improper dihedrals (°)	2.1

[a] $R\text{-merge} = \Sigma(I - \langle I \rangle))/\Sigma(I)$.
[b] $R\text{-factor} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$.

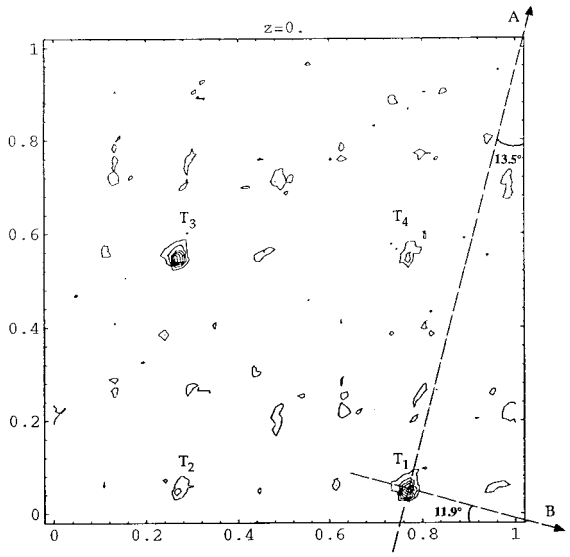

Figure 11. Translation function section of human pGSHPx at $Z = 0$. The cross-rotation solution after Patterson correlation refinement with Eulerian angles (θ_1, θ_2, θ_3) of (96.5°. 137.9°, 106.1°) was used in the translation search. This solution corresponds to the orientation of the local axis *B*. The translation search was performed in reciprocal space for the whole unit cell using data from 20 to 3.0 Å. The correct solution is the highest peak T_1 with fractional coordinates (X, Y) of (0.771, 0.048), which is well in line with the packing model (Ren *et al.*, 1995). T_1, T_2, T_3 and T_4, which could form a square with edges of 0.5 *a*, are equivalent peaks found in the translation search. The heights of T_1, T_2, T_3 and T_4 are 19.3 σ, 13.3 σ, 18.6 σ and 13.1 σ, respectively. The mean value of the translation function is 8.8 σ. This map is contoured from 8.8 σ in steps of 2.1 σ.

followed by 50 cycles of refinement with each subunit further divided into four rigid groups according to the secondary structure arrangements. During rigid-body refinement, the *R*-factor dropped to 40.5%. The free *R*-factor (Brünger, 1992b), for 10% of the observed diffraction data not used in refinement, was 44.7% at this stage.

An initial ($2F_o - F_c$) electron density map was calculated from the rigid-body refined model. This map showed clear density for a large proportion of residues. Density breaks, however, were observed in the regions of some loops, deletions, N and C terminals. This map was further improved by electron density averaging within the asymmetric unit.

Electron density averaging within a dimer

Real space electron density averaging was carried out with the program MAIN (Turk, 1992). The mask that covered the whole dimer was produced from the atomic coordinates by associating spheres with 4 Å radius around the atoms. The density map inside the mask was averaged about the non-crystallographic 2-fold axis (local axis *B*) using the operator obtained from least-squares superposition of the two subunits. The density averaging proceeded as a cyclic process: (1) two-fold averaging within the mask; (2) generation of the whole unit cell from the averaged map; (3) Fourier back transformation to obtain phases; and (4) calculation of a new electron density map for the next cycle. After the averaging process has converged, phase combination by Hendrickson-Lattman coefficients (Hendrickson & Lattman, 1970) was performed to combine the phases from the averaged map and the phases from the initial model. The combined phases were then used to calculate an electron density map that was used for model building. Generally, averaging improved the electron density for most of the residues. Only minor improvement was observed for surface residues such as the C-terminal residues. During model building, the averaged map was always used together with a ($2F_o - F_c$) map calculated directly with phases from the model.

Model building and crystallographic refinement

The molecular model was gradually improved by several cycles of model building and crystallographic refinement accompanied by electron density averaging. Model building was carried out using the program O (Jones *et al.*, 1991). In the first cycle, the real amino acid sequence was assigned to the 3D structure and the model was built for both subunits of the dimer independently. During model building, weights for atoms with absent or poor density were temporarily set to zero so that they would not contribute artificial phase information to the calculation of the electron density map for the next cycle. The model was refined with XPLOR using geometric parameters obtained from small-molecule data by Engh & Huber (1991). Reflection data between 8.0 and 2.9 Å was used in refinement. In the first cycle, the model was refined by conventional positional refinement followed by grouped *B*-factor and final positional refinements. Because of the limited resolution of the data, temperature factors were refined as two groups, one for main-chain atoms (CA, N, C and O) and one for side-chain atoms. In order to relax the model and improve the non-crystallographic symmetry (NCS) operator used in electron density averaging, NCS restraints were not applied to the refinement of the first model. The resulting model was subsequently used in electron density averaging and the second round of model building and refinement. The model was inspected on the graphics display and was extended by four N-terminal residues (6 to 9). This model was then subjected to molecular dynamics refinement and simulated annealing (Brünger, 1988) followed by grouped *B*-factor refinement. From the second cycle, NCS restraints were applied in all the following refinement cycles. Because different side-chain conformations were observed for the corresponding residues in the two subunits, looser restraints were put on side-chain atoms and tighter restraints on main-chain atoms (CA, N, C and O). The approximately optimal weights for NCS restraints, 300 kcal mol^{-1} $Å^{-2}$ for main-chain atoms and 5 kcal mol^{-1} $Å^{-2}$ for side-chain atoms, were determined by examination of the resulting model and the free *R*-factor with various weights. The simulated annealing (SA) refinement was performed from 3000 K to 300 K with a slow-cooling protocol. The SA refinement of the second model reduced the *R*-factor to 25.1% with the free *R*-factor of 36.7%. In subsequent cycles of model building and refinement, five more residues (1 to 5) were added to the N terminus and the rebuilt models were refined by positional and grouped *B*-factor refinements. Because the last 11 residues at the C terminus appeared only poorly defined in the electron density maps, they were not included in the final model. The final model, with 187 resi-

dues for each of the two subunits, was refined to an *R*-factor of 22.8% with a free *R*-factor of 33.5% for the resolution range of 8.0 to 2.9 Å. The refinement statistics are summarised in Table 4. Because of the limited resolution of the diffraction data, no attempt was made to add water molecules to the model. Throughout the model building and refinement process, the quality of the model was assessed by a number of criteria, including Ramachandran plots with the program PROCHECK (Laskowski *et al*., 1993); rotamer analysis (RSC_fit), peptide orientation analysis (Pep_flip) and residue real-space electron-density fit analysis (RS_fit) using the program O.

Acknowledgement

This work was supported by the Swedish Medical Research Council (grant no. B96-13X-09876-05A).

References

Avissar, N., Whitin, J. C., Allen, P. Z., Plamer, I. S. & Cohen, H. J. (1989a). Antihuman plasma glutathione peroxidase antibodies: immunologic investigation to determine plasma glutathione peroxidase protein and selenium content in plasma. *Blood,* **73**, 318–323.

Avissar, N., Whitin, J. C., Allen, P. Z., Wagner, D. D., Liegey, P. & Cohen, H. J. (1989b). Plasma selenium-dependent glutathione peroxidase, cell of origin and secretion. *J. Biol. Chem.* **264**, 15850–15855.

Avissar, N., Slemmon, J. R., Palmer, I. S. & Cohen, H. J. (1991). Partial sequence of human plasma glutathione peroxidase and immunologic identification of milk glutathione peroxidase as the plasma enzyme. *J. Nutr.* **121**, 1243–1249.

Avissar, N., Kerl, E. A., Baker, S. S. & Cohen, H. J. (1994). Extracellular glutathione peroxidase mRNA and protein in human cell lines. *Arch. Biochem. Biophys.* **309**, 239–246.

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

Björnstedt, M., Xue, J., Huang, W., Åkesson, B. & Holmgren, A. (1994). The thioredoxin and glutaredoxin systems are efficient electron donors to human plasma glutathione peroxidase. *J. Biol. Chem.* **269**, 29382–29384.

Broderick, D. J., Deagen, J. T. & Whanger, P. D. (1987). Properties of glutathione peroxidase isolated from human plasma. *J. Inorg. Biochem.* **30**, 299–308.

Brünger, A. T. (1988). Crystallographic refinement by simulated annealing. Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816.

Brünger, A. T. (1990). Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Crystallog. sect. A,* **46**, 46–57.

Brünger, A. T. (1992a). *XPLOR manual, version 3.1* Yale University, New Haven, CT.

Brünger, A. T. (1992b). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355**, 472–474.

Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W. & Harrison, P. R. (1986). The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the ''termination'' codon, TGA. *EMBO J.* **5**, 1221–1227.

Chou, P. Y. & Fasman, F. D. (1978). Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**, 251–276.

Chu, F.-F., Esworthy, R. S., Doroshow, J. H., Doan, K. & Liu, X.-F. (1992). Expression of plasma glutathione peroxidase in human liver in addition to kidney, heart, lung, and breast in humans and rodents. *Blood,* **79**, 3233–3238.

Chu, F.-F., Doroshow, J. H. & Esworthy, R. S. (1993). Expression, characterization, and tissue distribution of a new cellular selenium-dependent glutathione peroxidase, GSHPx-GI. *J. Biol. Chem.* **268**, 2571–2576.

Engh, R. & Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. *Acta Crystallog. sect. A,* **47**, 110–119.

Epp, O., Ladenstein, R. & Wendel, A. (1983). The refined structure of the selenoenzyme glutathione peroxidase at 0.2-nm resolution. *Eur. J. Biochem.* **133**, 51–69.

Esworthy, R. S., Chu, F.-F., Paxton, R., Akman, S. & Doroshow, J. (1991). Characterization and partial amino acid sequence of human plasma glutathione peroxidase. *Arch. Biochem. Biophys.* **286**, 330–336.

Esworthy, R. S., Chu, F.-F., Geiger, P., Girotti, A. W. & Doroshow, J. H. (1993). Reactivity of plasma glutathione peroxidase with hydroperoxide substrates and glutathione. *Arch. Biochem. Biophys.* **307**, 29–34.

Esworthy, R. S., Doan, K., Doroshow, J. H. & Chu, F.-F. (1994). Cloning and sequencing of the cDNA encoding a human testis phospholipid hydroperoxide glutathione peroxidase. *Gene,* **144**, 317–318.

Flohé, L. (1989). The selenoprotein glutathione peroxidase. In *Glutathione* (Dolphin, D., Avramovic, O. & Poulson, R., eds), pp. 644–731, John Wiley & Sons, New York.

Flohé, L., Loschen, G., Günzler, W. A. & Eichele, E. (1972). Glutathione peroxidase V. The kinetic mechanism. *Hoppe-Seyler's Z. Physiol. Chem.* **353**, 987–999.

Forstrom, J. W., Stults, F. H. & Tappel, A. L. (1978a). Rat liver cytosolic glutathione peroxidase: reactivity with linoleic acid hydroperoxide and cumene hydroperoxide. *Arch. Biochem. Biophys.* **193**, 51–55.

Forstrom, J. W., Zakowski, J. J. & Tappel, A. L. (1978b). Identification of the catalytic site of rat liver glutathione peroxidase as selenocysteine. *Biochemistry,* **17**, 2639–2644.

Gettins, P., Dyal, D. & Crews, B. (1992). Selenium-dependent glutathione peroxidase from ovine and bovine erythrocytes occur as longer chain forms than previously recognized. *Arch. Biochem. Biophys.* **294**, 511–518.

Günzler, W. A., Günzler, W. A., Steffens, G. J., Grossmann, A., Kim, S. A., Otting, F., Wendel, A. & Flohé, L. (1984). The amino-acid sequence of bovine glutathione peroxidase. *Hoppe-Seyler's Z. Physiol. Chem.* **365**, 195–212.

Hendrickson, W. A. & Lattman, E. E. (1970). Representation of phase probability distributions for simplified combination of independent phase information. *Acta Crystallog. sect. B,* **20**, 136–143.

Hendrickson, W., Horton, J. R. & LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672.

Huang, W. & Åkesson, B. (1993). Radioimmunoassay of glutathione peroxidase in human serum. *Clin. Chim. Acta,* **219**, 139–148.

Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF: a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in the model. *Acta Crystallog. sect. A,* **47**, 110–119.

Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallog.* **21**, 916–924.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopholymers,* **22**, 2577–2637.

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

Ladenstein, R., Epp, O., Bartels, K., Jones, A. & Huber, R. (1979). Structure analysis and molecular model of the selenoenzyme glutathione peroxidase at 2.8 Å resolution. *J. Mol. Biol.* **134**, 199–218.

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

Luzzati, P. V. (1952). Traitements statistiques des erreurs dans la determination des structures cristallines. *Acta Crystallog.* **5**, 802–810.

Maddipati, K. R. & Marnett, L. K. (1987). Characterization of the major hydroperoxide-reducing activity of human plasma: purification and properties of a selenium-dependent glutathione peroxidase. *J. Biol. Chem.* **262**, 17398–17403.

Maddipati, K. R., Gasparski, C. & Marnett, L. J. (1987). Characterization of the hydroperoxide-reducing activity of human plasma. *Arch. Biochem. Biophys.* **254**, 9–17.

Martin, J. L. (1995). Thioredoxin - a fold for all reasons. *Structure,* **3**, 245–250.

Martin-Alonso, J. M., Ghosh, S. & Coca-Prados, M. (1993). Cloning of the bovine plasma selenium-dependent glutathione peroxidase (GP) cDNA from the ocular cilliary epithelium: expression of the plasma and cellular forms within the mammalian eye. *J. Biochem.* **114**, 284–291.

Master, R. L., Magenheimer, B. S. & Calvet, J. P. (1994). Mouse plasma glutathione peroxidase. cDNA sequence analysis and renal proximal tubular expression and secretion. *J. Biol. Chem.* **269**, 27066–27073.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.

Mills, G. C. (1957). Hemoglobin catabolism. I. Glutathione peroxidase, an erythrocyte enzyme which protects hemoglobin from oxidative damage. *J. Biol. Chem.* **229**, 189–197.

Mullenbach, G. T., Tabrizi, A., Irvine, B. D., Bell, G. I., Tainer, J. A. & Hallewell, R. A. (1988). Selenocysteine's mechanism of incorporation and evolution revealed in cDNAs of three glutathione peroxidases. *Protein Eng.* **2**, 239–246.

Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Odom, J. D. (1983). Selenium biochemistry: chemical and physical studies. *Struct. Bond.* **54**, 1–26.

Reiter, R. & Wendel, A. (1983). Selenium and drug metabolism. I. Multiple modulations of mouse liver enzymes. *Biochem. Pharmacol.* **32**, 3063–3067.

Ren, B., Huang, W., Åkesson, B. & Ladenstein, R. (1995). Crystallization and preliminary X-ray diffraction analysis of glutathione peroxidase from human plasma. *Acta Crystallog. sect. D,* **51**, 824–826.

Richardson, J. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallog.* **15**, 24–31.

Schuckelt, R., Brigelius-Flohé, R., Maiorino, M., Roveri, A., Reumkens, J., Straβburger, W., Ursini, F., Wolf, B. & Flohé, L. (1991). Phospholipid hydroperoxide glutathione peroxidase is a selenoenzyme distinct from the classical glutathione peroxidase as evident from cDNA and amino acid sequencing. *Free Rad. Res. Commun.* **14**, 343–361.

Schwaab, V., Band, E., Ghyselinck, N., Mattei, M. G., Dufaure, J. P. & Drevet, J. R. (1995). Cloning of the mouse gene encoding plasma glutathione peroxidase: organization, sequence and chromosomal localization. *Gene,* **167**, 25–31.

Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G., Olin, B., Mannervik, B. & Jones, T. A. (1993). Structure determination and refinement of human alpha class glutathione transferase A1-1, and a comparison with the mu and pi class enzymes. *J. Mol. Biol.* **232**, 192–212.

Stadtman, T. C. (1987). Specific occurence of selenium in enzymes and amino acid tRNAs. *FASEB J.* **1**, 375–379.

Steigemann, W. (1974). Die Entwicklung und Anwendung von Rechenverfahren und Rechen programmen zur Strukturanalyse von Proteinen au Beispiel des Trypsin-Trypsininhibitor-komplexes, des freien Inhibitors und der L-Asparaginase. Doctoral thesis, Technical University of Munich, Germany.

Takahashi, K., Avissar, N., Whitin, J. & Cohen, H. J. (1987). Purification and characterization of human plasma glutathione peroxidase: a selenoglycoprotein distinct from the known cellular enzyme. *Arch. Biochem. Biophys.* **256**, 677–686.

Takahashi, K., Akasaka, M., Yamamoto, Y., Kobayashi, C., Mizoguchi, J. & Koyama, J. (1990). Primary structure of human plasma glutathione peroxidase deduced from cDNA sequences. *J. Biochem.* **108**, 145–148.

Thomas, J. P., Maiorino, M., Ursini, F. & Girotti, A. W. (1990). Protective action of phospholipid hydroperoxide glutathione peroxidase against membrane-damaging lipid peroxidation: in situ reduction of phospholipid and chlesterol hydroperoxides. *J. Biol. Chem.* **265**, 454–461.

Turk, D. (1992). Development and usage of a molecular graphics program. Doctoral thesis, Technical University of Munich, Germany.

Ursini, F., Maiorino, M., Brigelius-Flohé, R., Aumann, K. D., Roveri, A., Schomburg, D. & Flohé, L. (1995). Diversity of glutathione peroxidases. *Methods Enzymol.* **252**, 38–53.

Venkatachalam, C. M. (1968). Stereochemical criteria for polypeptides and proteins V. Conformation of a

system of three linked peptide units. *Biopolymers,* **6**, 1425–1436.

Von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucl. Acids Res.* **14**, 4683–4690.

Wilson, A. J. (1949). The probability distribution of X-ray intensities. *Acta Crystallog.* **2**, 318–321.

Yamamoto, Y. & Takahashi, K. (1993). Glutathione peroxidase isolated from plasma reduces phospholipid hydroperoxides. *Arch. Biochem. Biophys.* **305**, 541–545.

Yoshimura, S., Watanabe, K., Suemizu, H., Onozawa, T., Mizoguchi, J., Tsuda, K., Hatta, H. & Moriuchi, T. (1991). Tissue-specific expression of the plasma glutathione peroxidase gene in rat kidney. *J. Biochem.* **109**, 918–923.

Zinoni, F., Birkmann, A., Stadtman, T. C. & Böck, A. (1986). Nucleotide sequence and expression of selenocysteine-containing polypeptide of formate dehydrogenase (format-hydrogen-lyase-linked) from *Escherichia coli. Proc. Natl Acad. Sci. USA,* **83**, 4650–4654.

Edited by R. Huber

(Received 21 November 1996; received in revised form 17 February 1997; accepted 20 February 1997)