Structure, Vol. 9, 1117–1125, November, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00666-9

# *Escherichia coli* GlpE Is a Prototype Sulfurtransferase for the Single-Domain Rhodanese Homology Superfamily

**Andrea Spallarossa,[1] Janet L. Donahue,[2] Timothy J. Larson,[2] Martino Bolognesi,[3] and Domenico Bordo[4,5]**

[1] Department of Pharmaceutical Sciences
University of Genova
Viale Benedetto XV, 3
16132 Genova
Italy

[2] Department of Biochemistry
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

[3] Department of Physics
INFM
University of Genova c/o Advanced Biotechnology Center
Largo R. Benzi 10
16132 Genova
Italy

[4] National Institute for Cancer Research
c/o Advanced Biotechnology Center
Largo R. Benzi 10
16132 Genova
Italy

## Summary

**Background:** Rhodanese domains are structural modules occurring in the three major evolutionary phyla. They are found as single-domain proteins, as tandemly repeated modules in which the C-terminal domain only bears the properly structured active site, or as members of multidomain proteins. Although in vitro assays show sulfurtransferase or phosphatase activity associated with rhodanese or rhodanese-like domains, specific biological roles for most members of this homology superfamily have not been established.

**Results:** Eight ORFs coding for proteins consisting of (or containing) a rhodanese domain bearing the potentially catalytic Cys have been identified in the *Escherichia coli* K-12 genome. One of these codes for the 12-kDa protein GlpE, a member of the *sn*-glycerol 3-phosphate (*glp*) regulon. The crystal structure of GlpE, reported here at 1.06 Å resolution, displays $\alpha/\beta$ topology based on five $\beta$ strands and five $\alpha$ helices. The GlpE catalytic Cys residue is persulfurated and enclosed in a structurally conserved 5-residue loop in a region of positive electrostatic field.

**Conclusions:** Relative to the two-domain rhodanese enzymes of known three-dimensional structure, GlpE displays substantial shortening of loops connecting $\alpha$ helices and $\beta$ sheets, resulting in radical conformational changes surrounding the active site. As a consequence, GlpE is structurally more similar to Cdc25 phosphatases than to bovine or *Azotobacter vinelandii* rhodaneses. Sequence searches through completed genomes indicate that GlpE can be considered to be the prototype structure for the ubiquitous single-domain rhodanese module.

## Introduction

Sulfurtransferases (E.C. 2.8.1.x) are widespread enzymes, found in all major evolutionary phyla, yet their biological role is largely debated. The thiosulfate:cyanide sulfurtransferases, or rhodaneses (E.C. 2.8.1.1), catalyze the transfer of a sulfane sulfur atom from thiosulfate to cyanide; the 3-mercaptopyruvate sulfurtransferases (MST, E.C. 2.8.1.2) catalyze the same sulfane sulfur transfer reaction, using 3-mercaptopyruvate as a sulfur donor [1]. Rhodaneses and MSTs display evident amino acid sequence homology (about 40% identical residues), indicative of a common evolutionary origin and similar three-dimensional folds [2]. The best-characterized sulfurtransferase is bovine rhodanese (Rhobov), which catalyzes the formation of sulfite and thiocyanate from thiosulfate and cyanide [3, 4]. In the course of catalysis, the enzyme cycles between a sulfur-free form and a persulfurated intermediate, hosting the persulfide sulfur atom on the catalytic Cys residue. Given its abundance in bovine liver mitochondria, Rhobov has been deemed as primarily involved in cyanide detoxification [1]. However, other functional roles for rhodanese enzymes have been proposed, including sulfur insertase in the biosynthesis and/or repair of iron-sulfur clusters, involvement in sulfur metabolism, or interaction with thioredoxin [4–8].

The three-dimensional structures of two known sulfurtransferases, Rhobov and *Azotobacter vinelandii* rhodanese (RhdA) [9, 10], have shown that, in spite of limited sequence similarity (22% identical residues), both enzymes are composed of two structurally similar domains that may have originated by ancestral gene duplication [9]. These domains, henceforth called rhodanese domains, both display a typical $\alpha/\beta$ topology. The essential Cys residue, located in the C-terminal domain of both enzymes, is the first residue of a 5 amino acid loop (the active-site loop) that folds in a cradle-like structure defining the enzyme catalytic pocket. The catalytic Cys thiol is located at the bottom of the pocket, within a strongly positive electrostatic field generated by the architecture of the active site and by the surrounding basic residues [11]. In spite of the similar thiosulfate:cyanide sulfurtransferase activity and the conserved catalytic Cys residue, the amino acid composition and the location of charged residues in the active-site region of Rhobov and RhdA are widely distinct [10].

Recently, the rhodanese fold has been unexpectedly observed in the structure of the catalytic domains of the human cell cycle-control phosphatases Cdc25A and

[5] Correspondence: domenico@alcor.ge.infm.it

**Key words:** Cdc25; phosphatase; rhodanese; sulfurtransferase

Cdc25B [12, 13], which host the active-site Cys residue at a strictly comparable site, despite the negligible amino acid sequence similarity to the rhodanese family. Moreover, genome sequencing has shown that ORFs coding for rhodanese (or MST) homologs are present in most eubacteria, archea, and eukaryota (see, e.g., COG [14] or SMART [15]). Often, several genes coding for distinct "rhodanese-like" proteins are found in the same genome, suggesting that the encoded proteins may have distinct biological functions.

From the structural viewpoint, rhodanese-like proteins are either found in the usual tandem domain arrangement or are composed of a single rhodanese domain. Examples of single rhodanese domain proteins are the sulfide dehydrogenase (Sud) from *Wolinella succinogenes* [16] and the dark-inducible, senescence-associated protein Ntdin from *Nicotiana tabacum* (EMBL accession number, Q9MBD6). Finally, rhodanese domains are also found as modules in multidomain proteins displaying widely different enzymatic activities. For example, a rhodanese domain is found in the C-terminal part of ThiI, a 482 amino acid enzyme from *Escherichia coli* involved in the biosynthesis of thiamin and 4-thiouridine [17], and in the NADH oxidase (NoxA-3) from *Archaeoglobus fulgidus*.

The *E. coli* genome [18] encodes eight proteins containing rhodanese domains with the putative active-site Cys residue at the expected position. One of them is GlpE, a 12-kDa protein that displays weak but significant sequence similarity (17% identical residues) to the C-terminal domain of RhdA, with a Cys residue (Cys65) matching the expected sulfurtransferase active-site location [8]. Biochemical characterization in vitro has shown that GlpE may act as a rhodanese catalyzing, although with low efficiency, the sulfur transfer reaction from thiosulfate to cyanide [8]. To shed more light on the function of GlpE, in the context of a wider investigation of the biological role and structural adaptability of rhodanese modules, we determined the three-dimensional structure of GlpE (at 1.06 Å resolution) and that of its intermediate in the sulfurtransferase catalytic cycle. The structural properties of GlpE are compared with those of homologous sulfurtransferases and are discussed in light of multiple rhodanese domains encoded by the genomes of *E. coli* and other organisms.

## Results

### Overall Three-Dimensional Structure of GlpE

The 1.06 Å crystal structure of GlpE was solved using the SIRAS (single isomorphous replacement with anomalous scattering) method on a single heavy atom derivative obtained by soaking GlpE native crystals in $HoSO_4$ solutions, as described in the Experimental Procedures. The initial phases, calculated at 2.0 Å resolution, were improved by solvent flattening and phase extension to the resolution limit of 1.06 Å. The GlpE model (108 amino acids) could then be entirely built in the experimental electron density map, which resulted in outstanding quality, and could be refined to a final R factor of 12.8% ($R_{free}$ = 15.1%). The final GlpE model (847 protein atoms) contains 97 water molecules, 7 ethanediol molecules (part of the cryoprotectant solution), and 2 acetate anions (see Table 1).

A single GlpE molecule is observed in the crystallographic asymmetric unit of the trigonal crystal form analyzed. In the crystalline lattice, each GlpE molecule establishes direct contacts with eight neighbors. Intermolecular interactions involve an interface of, at most, 458 Å$^2$, for a total of 2237 Å$^2$, representing 37% of the total GlpE surface (5980 Å$^2$). Intermolecular interactions are mostly based on van der Waals contacts, with only four hydrogen bonds. Although GlpE is reported to adopt a dimeric assembly under native gel chromatography conditions [8], analysis of the molecular packing does not allow identification of a firm dimeric assembly for the crystallized enzyme. More specifically, crystalline GlpE does not mimic the (covalent) tandem-domain arrangement observed in RhdA and in Rhobov, in which the domain interface covers about 2200 Å$^2$ [10].

As anticipated by the weak but significant amino acid sequence homology to other rhodanese enzymes, GlpE adopts the three-dimensional fold typical of a single $\alpha/\beta$ rhodanese domain (see Figures 1 and 4a). Such structural similarity, however, was not sufficient to allow solution of the GlpE structure via molecular replacement techniques. The three-dimensional structure of GlpE is based on a central parallel $\beta$ sheet composed of five $\beta$ strands, in the order 1, 5, 4, 2, and 3 ($\beta$A, $\beta$E, $\beta$D, $\beta$B, and $\beta$C in Figure 1b), arranged in the conventional counterclockwise twist. Two and three $\alpha$ helices pack against each side of the $\beta$ sheet, respectively. A second antiparallel, small $\beta$ sheet, formed by residues 36–37 and 104–105, may be involved in structural stabilization of the protein's C-terminal segment.

### GlpE Active Site

The active-site residue Cys65 is the first amino acid of the active-site loop connecting the central $\beta$ strand ($\beta$D, residues 61–64) of the 5-stranded $\beta$ sheet to the 70–81 $\alpha$D helix. The active-site loop (-Cys65-Tyr66-His67-Gly68-Asn69-Ser70-, Figure 2a) forms a cradle-like, semicircular structure centered around the Cys65 S$\gamma$ atom. Native GlpE crystals display a strong electron density peak, 2.11 Å from the Cys65 S$\gamma$ atom (Figure 2a). For this density feature, anomalous Fourier difference maps (at 0.990 Å X-ray wavelength) showed an anomalous signal (6.0 $\sigma$) very similar to that of the S$\gamma$ atom of Cys65 (7.5 $\sigma$). Such an observation, together with the observed bond length, strongly suggests the presence of an extra sulfur atom (S$\delta$) involved in a persulfide bond with the Cys65 S$\gamma$ atom. In native GlpE, the persulfide S$\delta$ atom (B factor, 12.6 Å$^2$) is at the center of a hydrogen bond radial network built by the five main chain amide groups of the active-site loop and by the hydroxyl group of Ser70 (Figure 2a). The S$\delta$ atom may also be involved in hydrogen bonds with two water molecules located in front of the active site (S$\delta$-O distances, 3.34 Å and 3.35 Å; Figure 2a).

Native GlpE crystals also display electron density corresponding to an alternate conformation for Cys65, characterized by a 110° rotation of the thiol group about the C$\alpha$-C$\beta$ axis. In this minor conformer, Cys65 does not bear the persulfide sulfur, the S$\gamma$ atom being located

Table 1. Crystal Structure Analysis

| | Native | Holmium Derivative | KCN Soak |
|---|---|---|---|
| **Data Collection Statistics** | | | |
| Resolution limit (Å) | 1.06 | 2.0 | 1.8 |
| X-ray source | synchrotron (0.910 Å) | synchrotron (0.939 Å) | CuK$\alpha$ (1.54 Å) |
| Mosaicity (°) | 0.45 | 0.5 | 0.7 |
| Number of observations | 555,366 | 55,737 | 64,419 |
| Unique reflections | 45,040 | 6,806 | 9,094 |
| Completeness (%) | 99.6 | 97.0 | 95.7 |
| I/$\sigma$(I) | 42.7 | 15.9 | 22.7 |
| $R_{sym}$ (%)[a] | 2.3 | 7.8 | 3.2 |
| Outer resolution shell completeness (%) | 93.6 | 74.0 | 84.7 |
| Outer resolution shell $R_{sym}$ (%)[a] | 28.1 | 8.1 | 23.4 |
| Phasing power[b] | | 2.2 | |
| Anomalous phasing power[c] | | 4.3 | |
| Overall figure of merit | | 0.72 | |
| **Refinement Statistics** | | | |
| Nonhydrogen atoms | | | |
| Protein | 847 | | 846 |
| Solvent | 106 | | 82 |
| R factor (%)[d] | 12.8 | | 17.7 |
| $R_{free}$ (%)[e] | 15.1 | | 24.8 |
| Rmsd | | | |
| Bond distances (Å) | 0.018 | | 0.013 |
| Bond angle distances (°) | 2.3 | | 1.4 |
| Average B factor | | | |
| Main chain atoms (Å$^2$) | 13.5 | | 24.7 |
| Side chain atoms (Å$^2$) | 17.4 | | 28.6 |

[a] $R_{sym} = \Sigma|(I - <I>)|/\Sigma(I)$, where I is the observed intensity.
[b] Phasing power = $\Sigma|F_H|/\Sigma||F_{PHobs}| - |F_{PHcalc}||$.
[c] Anomalous phasing power = $\Sigma|F_H$ (imag)$|/\Sigma||F_{PHobs}| - |F_{PHcalc}||$.
[d] R factor = $\Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$.
[e] $R_{free}$ was calculated on 5% randomly selected reflections, which were not included in the refinement.

on a side of the catalytic pocket, hydrogen bonded to the Ser70 O$\gamma$ atom. The occupancies of the first (persulfurated) and the second (persulfide free) Cys65 conformers were evaluated as 80% and 20%, respectively. The presence of two conformational states may be related to the purification protocol adopted for GlpE, in which two distinct peaks obtained from anion-exchange chromatography, possibly corresponding to distinct active-site conformation/ionization states of GlpE, were pooled together for crystallization purposes (see the Experimental Procedures).

The active-site loop main chain conformation closely matches that observed in RhdA and Rhobov (C$\alpha$-C$\alpha$ rmsd calculated over the five residues of the active-site loop, 0.32 Å and 0.36 Å, respectively), in spite of the distinct amino acid composition (Figure 3). Five peptide dipoles provided by amino acids 66–70 aim at the center of the catalytic pocket, leading to the establishment of a positive electrostatic field surrounding the catalytic residue. The estimated strength of the electrostatic field, calculated by using a linearized Poisson-Boltzman equation as implemented in DELPHI [19], reaches the value of 13 *kT/e*, a value similar to that calculated for RhdA and Rhobov (18 *kT/e* and 19 *kT/e*, respectively). These findings indicate that the strength and shape of the electrostatic field is primarily dependent on the main chain conformation of the active-site loop, rather than on the amino acid composition.

### Structural Bases of the Sulfurtransferase Activity

To confirm on a structural basis the in vitro-observed thiosulfate:cyanide sulfutransferase activity [8], native (i.e., 80% persulfurated) GlpE crystals were treated with cyanide, as described in the Experimental Procedures (Table 1), and the resulting crystal structure was analyzed at 2.0 Å resolution (R factor = 17.7%, $R_{free}$ = 24.8%). As expected, cyanide soaking led to removal of the Cys65 persulfide S$\delta$ atom. The Cys65 side chain was still observed as a mixture of two alternate conformations, close to those observed in native GlpE crystals (Figure 2b), each with a relative occupancy of 50%. An electron density peak at the center of the catalytic pocket falls at 2.77 Å from the Cys65 S$\gamma$ atom. Such contact distance is compatible with the presence of a $Na^+$ ion at this position, stabilized by coulombic interaction with Cys65 thiolate. Moreover, the Cys65 S$\gamma$ atom is at hydrogen bonding distance from the main chain amide groups of the residues that form the active-site loop: Try66 (3.33 Å), Gly68 (3.42 Å), Asn69 (3.60 Å), and Ser70 (3.55 Å), as well as from the Ser70 O $\gamma$ atom (3.28 Å), where the S$\gamma$-N distance is given within parentheses.

Following the observation that, in RhdA, the active-site Cys persulfide could be cleaved by phosphor-containing compounds [20], native GlpE crystals were soaked in $K_2HPO_4$ and in $H_3PO_2$ solutions. Crystallographic analysis showed that neither compound was able to remove the Cys65 persulfide sulfur or to affect the active-site conformation. This agrees with the observation by Ray

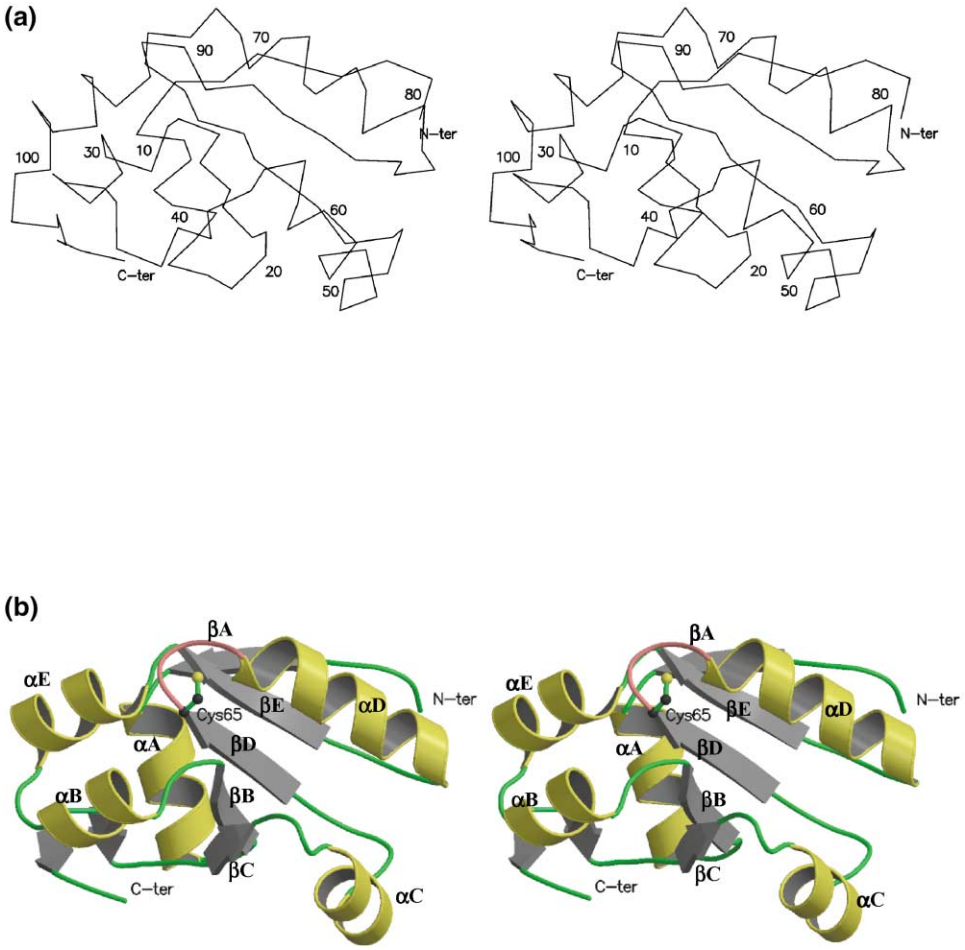


Figure 1. Stereo Representation of GlpE Three-Dimensional Structure

(a) Cα trace. Every tenth residue and the N- and C-terminal residues are labeled.

(b) Ribbon diagram. Labels for secondary structure elements forming the α/β rhodanese fold are the same as those previously adopted for Rhobov and RhdA [9, 10]. The active-site loop is shown in pink, with the active-site Cys65 represented in ball-and-stick. All drawings were prepared with MOLSCRIPT [50] and Raster3D [51].

et al. [8] that there are no phosphate inhibitory effects on the sulfurtransferase activity of GlpE.

### Comparison with Rhodaneses and Cdc25 Phosphatase Enzymes: Conserved Structural Signatures

As described above, the three-dimensional structure of GlpE is clearly related to that of the C-terminal (catalytic) domain of the other two rhodanese enzymes of known structure, Rhobov or RhdA [9, 10]. However, the connecting loops within the α/β rhodanese domain are shortened in GlpE to a minimal length, resulting in a 108-residue protein, while the C-terminal domains of RhdA and Rhobov are composed of 127 and 136 amino acids, respectively. In particular, the linker peptide connecting αB to βC spans 18 residues in Rhobov, where it folds in front of the active-site Cys247; this loop is 12 amino acids shorter in GlpE (Figures 3 and 4a). Moreover, the structural comparisons highlight the role of the βC-αC linker in stabilizing the tandem domain arrangement of RhdA and Rhobov where, due to the interdomain pseudo two-fold axis, the linker peptide folds (in both domains) against the other rhodanese domain (Figure 4a). In GlpE, instead, the minimal size of αB-βC and βC-αC linkers results in a substantial structural modification of the surface region surrounding the catalytic pocket, which falls on a convex region of GlpE, as opposed to a depression in the protein surface of RhdA or Rhobov.

Despite the above considerations, the active-site main chain conformation is strongly preserved in GlpE, RhdA, and Rhobov. In spite of the observed amino acid variability (Figure 3), the active-site loop conformation in the three proteins is such that only the side chain of the residue at position Cys+3 points away from the active site, whereas the other four side chains are oriented to putatively interact with the substrate (see Figures 2a and 2b). At least two polar (often charged) residues are observed at these four sites in rhodanese enzymes. Such a conserved trend is likely related to substrate binding/selectivity and/or to the stabilization of the persulfurated form of the enzyme.

Notably, structural comparison with the catalytic subunit of the human cell cycle phosphatase Cdc25A [12]

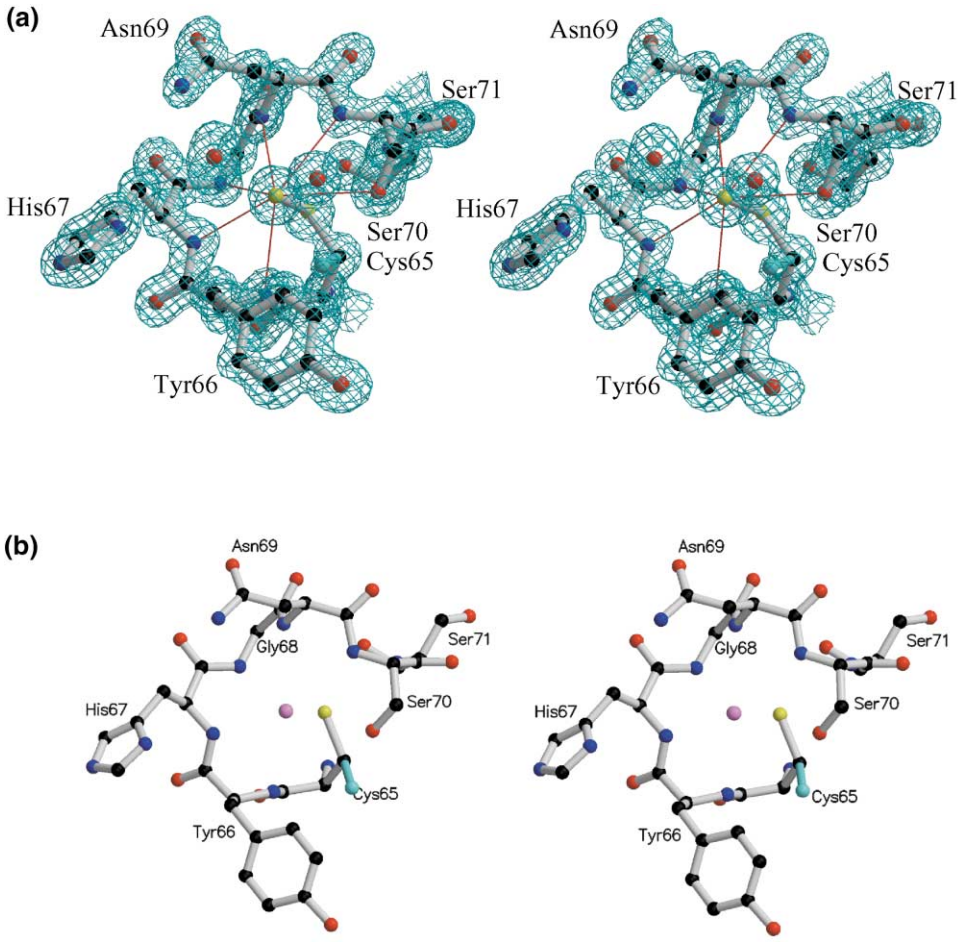


Figure 2. Stereo Representation of the GlpE Active-Site Loop

(a) The active site as observed in native GlpE, with the persulfide sulfur atom (Sδ) bound to the Cys65 Sγ atom. The electron density map ($2F_o$-$F_c$) is contoured at 1.0 σ. Hydrogen bonds involving Sδ in its major conformer are shown as red lines. Hydrogen bonds involving the two water molecules observed in front of the active site are not drawn for clarity. O, N, C, and S atoms are represented in red, blue, black, and yellow, respectively. The two alternate Cys65 conformations are shown in gray and light blue, respectively. The drawing was prepared with BOBSCRIPT [52].

(b) Sulfur-free GlpE after soaking with KCN, showing a sodium cation (in magenta) at the center of the active site, bound to the active-site Cys65 thiol group.

shows that the overall structural similarity between GlpE and Cdc25A (rmsd of 1.80 Å, calculated over 96 equivalent Cα pairs) is higher than for GlpE-RhdA (rmsd of 2.14 Å, calculated over 101 Cα pairs) or GlpE-Rhobov (rmsd of 2.13 Å, calculated over 97 Cα pairs). Similar results are obtained when GlpE is compared with Cdc25B [13]. In particular, the αB-βC and the βC-αC linker peptides match quite well in GlpE and Cdc25A (Figures 3 and 4b), resulting in very similar main chain conformations surrounding both catalytic centers. Conversely, the opposite faces of each protein (relative to the active site) display substantial structural deviations, as indicated by the αA-βB, αC-βD, and αD-βE loops (Figures 3 and 4b), which are longer in Cdc25 enzymes, possibly related to the association with the N-terminal (regulatory) Cdc25 domain (full-length Cdc25A and Cdc25B are 523 and 566 amino acids long, respectively).

The structure-based sequence alignment of known rhodanese modules GlpE, RhdA, and Rhobov and of Cdc25 phophatases (Figure 3) allows for identification of conserved amino acids that may define signatures of functional and/or structural importance. Apart from the active-site Cys residues, two other amino acids, Asp25 and Arg27 (GlpE numbering is used), are conserved in these enzymes (Figure 3) and in most other rhodanese domains [10]. In all known rhodanese and Cdc25 phosphatase structures, Asp25 and Arg27 form a hydrogen-bonded salt bridge. As Asp25 is totally buried in the protein core and the guanidinium group of Arg27 is only minimally exposed to the solvent (14 $Å^2$), the contribution of this interaction to the overall free energy of the rhodanese-like domain must be substantial.

## Discussion

The model of GlpE described herein provides the first three-dimensional structure of an enzyme acting in vitro as a sulfurtransferase, based on a single rhodanese domain. Comparison of GlpE with the catalytic domain of Rhobov and RhdA highlights a substantial editing of

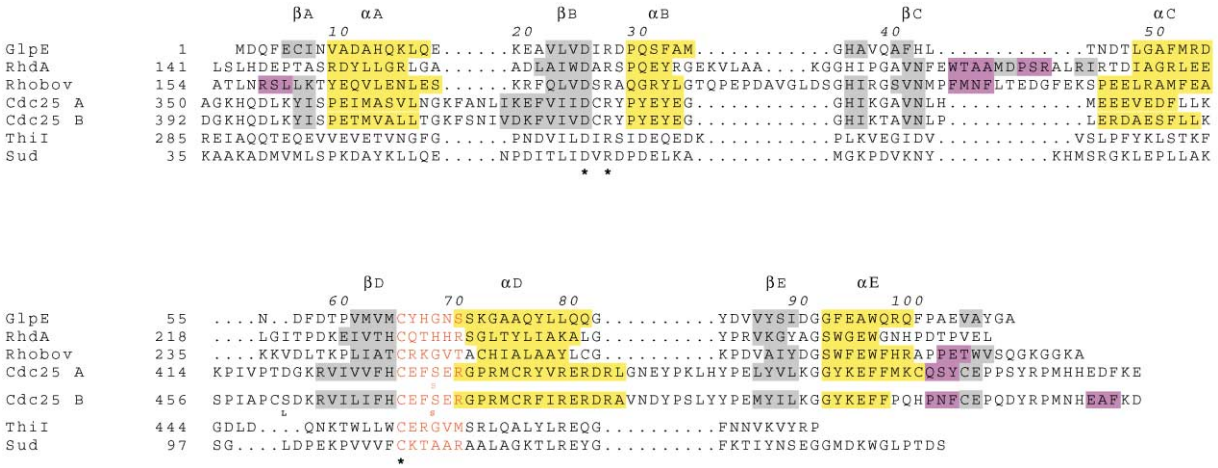


Figure 3. Multiple Sequence Alignment of the Rhodanese Structures and of Cdc25A and Cdc25B Phosphatases

The sequence alignment was obtained by optimal structural superposition on GlpE, taken as reference, of the C-terminal domains of RhdA and Rhobov, and the catalytic domain of Cdc25A and Cdc25B phosphatases. Residues assuming α-helical, $3_{10}$-helical, or β strand conformation are included in yellow, magenta, and gray boxes, respectively. Labels for α helices and β strands defining the α/β rhodanese fold are reported on the first line. Active-site loop residues are shown in red. The amino acid sequence of ThiI and Sud, for which sulfurtransferase activity has been reported, are also shown. Three single amino acid insertions in the Cdc25A and Cdc25B sequences have been represented as small characters below the respective sequence at the corresponding position. Asterisks denote invariant residues.

the loops connecting the secondary structure elements, which include 39% of GlpE residues, as opposed to 62% and 57% in RhdA and Rhobov, respectively.

The GlpE active-site loop is structurally reminiscent of the phosphate binding loop observed in members of the protein tyrosine phosphatase and Cdc25 phosphatase protein families. The loops in the phosphatases, however, are one residue longer. The peculiar active-site loop conformation contributes to the buildup of a positive electrostatic field, expected to lower the $pK_a$ of the catalytic Cys residue to values as low as 6.5 in rhodanese enzymes [21–23]. A similar $pK_a$ shift for the catalytic Cys residue has been observed in *Yersinia* tyrosine phosphatase, which displays a similar active-site loop structure [24].

GlpE, as well as RhdA, displays a stable persulfurated native enzyme form [10, 20]. Stabilization of the persulfurated catalytic Cys is supported by an extensive hydrogen bonding network radially centered on the persulfide sulfur, reflecting a precise tailoring (and likely substrate selectivity) of the catalytic pocket. The lack of inhibitory or structural effects of phosphate or hypophosphite on GlpE [8] does not match the behavior of RhdA, where these compounds remove the persulfide sulfur [20]. The latter property may be linked to the presence of Arg235 as the last residue in the RhdA active-site loop (Ser70 in GlpE). Arg is also found in this position in Cdc25 phosphatases and likely provides stabilization for an active-site anion.

The observed abundance of potentially functional rhodanese-like proteins in the *E. coli* genome as well as in other genomes, together with the amino acid variability observed around the catalytic Cys residue, strongly suggests that members of this homology superfamily may play distinct biological roles. In vitro assays have shown that GlpE, RhdA, *E. coli* ThiI, and *Arabidopsis thaliana* AtRDH1 and AtRDH2 display thiosulfate:cyanide sulfurtransferase activities [8, 17, 25, 26]. *Wolinella succinogenes* Sud has been recently characterized as a polysulfide:cyanide sulfurtransferase [27], while biochemically characterized MSTs include proteins from rat, *A. thaliana*, and *E. coli* [28–31]. Comparison of their in vitro activities reveals quite distinct catalytic efficiencies. For instance, GlpE has rather high $K_m$ values for both thiosulfate and cyanide, and the enzyme is inactivated upon prolonged treatment with thiosulfate [8]. These observations raise questions about the actual physiological significance of these compounds as GlpE substrates. Moreover, the observed ability of phosphate- and hypophosphite-containing compounds to cleave the active-site persulfide in RhdA suggests that sulfurtransfer to cyanide may not represent the primary biological function, at least for some rhodanese-like proteins [20].

Despite the increasing amount of structural information, there remains the open question of the identification of the true biological substrates, either proteins or small molecules, that donate and remove the persulfide sulfur from the active-site Cys residue. As is the case for Cdc25 phosphatases, substrate selectivity may be very tight.

## Biological Implications

The complete sequencing of a variety of genomes has shown that multiple ORFs coding for rhodanese-like proteins are often present in the same genome. In particular, the *E. coli* genome contains eight ORFs coding for proteins composed of at least one rhodanese domain, with a putatively catalytic Cys at the expected position. Three of them code for proteins GlpE, PspE, and YgaP, which are comprised of a single rhodanese domain. The GlpE three-dimensional structure has been shown here to match that of the catalytic domain of prokaryotic and eukaryotic tandem-domain rhodaneses. Two other

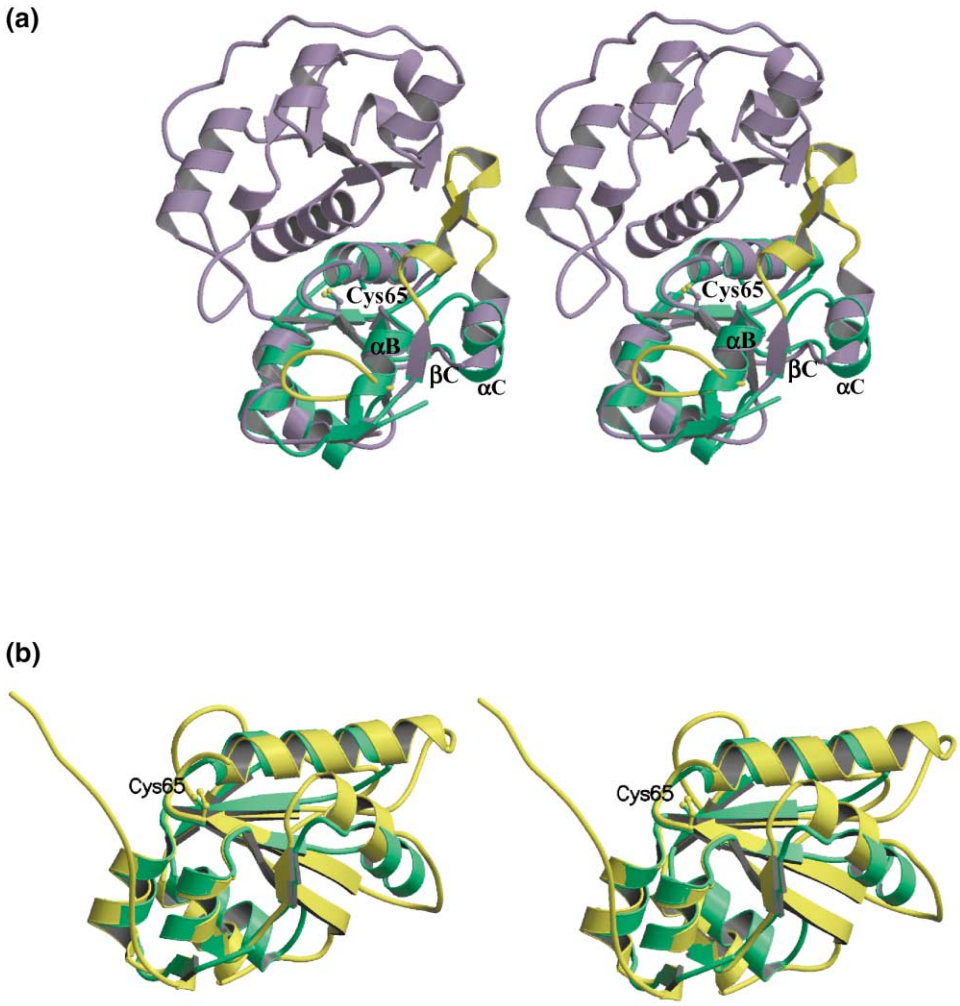


Figure 4. Structural Comparisons between GlpE, RhdA, and Cdc25A Phosphatase

(a) GlpE versus RhdA C-terminal domain. GlpE and RhdA molecules are shown in green and pink, respectively. RhdA αB-βC and βC-αC linkers are represented in yellow. The catalytic Cys residues are represented in ball-and-stick.

(b) GlpE versus Cdc25A catalytic domain; the two enzymes are shown in green and blue, respectively.

proteins, SseA and YnjE, are composed of two rhodanese domains, with the C-terminal domain containing the putatively catalytic Cys. Three ORFs code for proteins ThiI (482 amino acids), YbbB (364 amino acids), and YceA (350 amino acids) that contain a single rhodanese module in the context of a larger protein construct. Sulfurtransferase activity has been previously described for ThiI and SseA. The amino acid variability observed for the putative active-site loops in the identified homologs suggests diversification of substrate specificity, while maintaining enzymatic activities related to the formation, interconversion, and transport of compounds containing sulfane sulfur atoms. The significant sequence similarity that relates GlpE to the periplasmic sulphide dehydrogenase (Sud) from *W. succinogenes*, for which sulfurtransferase activity has been described recently, as well as the strong structural similarity to Cdc25 phosphatases further points to GlpE as the prototypical structure for single-domain rhodanese-like enzymes. Genetic techniques, such as the generation of multiple knockout mutants, as well as the use of bioinformatic approaches based on recurring instances of neighboring genes in distinct genomes, will likely provide a more definite answer to the biological roles and substrates of this ubiquitous protein superfamily.

## Experimental Procedures

### Protein Expression, Purification, and Crystallization

GlpE was subcloned and overexpressed in *E. coli* strain BL21(DE3)(pMS421), following the previously described protocol, with minor modifications [8, 32]. Protein purification was performed by freeze-thaw treatment, followed by anion-exchange chromatography. Two partially overlapping peaks, eluted at 160 mM and 180 mM NaCl, contained most of the rhodanese activity and exhibited close electrophoretic mobility. For crystallization purposes, the two peaks were pooled and concentrated by step elution chromatography. The final GlpE preparation displayed no detectable impurities as assessed by Coomassie blue staining (15% SDS polyacrylamide gel, 3 μg GlpE per lane). Rhodanese activity of purified GlpE, measured according to the conventional Sorbo assay [3], was ~200 U $mg^{-1}$, where 1 U is the amount of enzyme catalyzing the transforma-

tion of 1 μmol of substrate per minute. For crystallization purposes, the purified protein was brought to a final concentration of 8 mg $ml^{-1}$, in the presence of 50 mM Tris/HCl (pH 7.0), 3 mM EDTA, 100 mM NaCl, and 10% (v/v) glycerol. Crystals of optimal quality were obtained with the hanging drop setup. The precipitant solution contained 0.1 M $CaCl_2$, 0.1 M $CH_3COONa$ (pH 4.6), and 20% (v/v) 2-propanol, as described previously [32].

### Data Collection and Structure Determination

GlpE crystals displayed excellent X-ray diffraction power. Synchrotron data collection, performed at EMBL c/o DESY (Hamburg), could be carried out to a resolution limit of 1.06 Å on the native protein (Table 1). Crystals belong to the trigonal space group $P3_2$ ($a = b = 53.87$ Å, $c = 30.52$ Å, $\gamma = 120°$). Evaluation of the crystal packing parameters [33] indicates the presence of one molecule per asymmetric unit and a rather low solvent content (42% v/v). Data were reduced and scaled with DENZO and SCALEPACK, respectively [34, 35]. Further data analysis was carried out with programs of the CCP4 program suite [36].

In spite of the weak but significant sequence similarity of GlpE with the C-terminal domain of RhdA (17% identical residues), molecular replacement [37, 38] using several search models derived from RhdA provided no solution to correctly locate the template molecule within the GlpE unit cell. Structure solution was therefore pursued using one heavy atom isomorphous derivative obtained by soaking GlpE native crystals in a $HoSO_4$ solution (5 mM) for a period of 12 hr. A derivative dataset was collected at ESRF-Grenoble, at 0.939 Å wavelength, to a resolution limit of 2.0 Å (Table 1). Both isomorphous and anomalous Patterson maps displayed a single heavy atom site, which was used to obtain initial phases. For this purpose, PHASES [39] was used. The heavy atom parameters were further refined using the maximum likelihood technique as implemented in SHARP [40]. Phases, initially calculated at 2.0 Å resolution, were gradually extended to the 1.06 Å resolution limit with SOLOMON [41]. The final solvent-flattened electron density map was of outstanding quality, with the whole GlpE molecule clearly defined.

Data were collected on native GlpE crystals soaked with KCN, $K_2HPO_4$, and $H_3PO_2$ (5 mM each) using a conventional rotating anode X-ray generator, at the Cu Kα wavelength.

All data collection was carried out at 100 K. For this purpose, GlpE crystals were transferred in a cryoprotectant solution identical to the crystallization medium but containing an additional 20% (v/v) ethanediol.

### Model Refinement and Analysis

WARP [42] was used to build the whole protein backbone and the side chains in an automated way. The protein model was refined with REFMAC [43]. After conventional refinement of atomic coordinates with isotropic B factors, water, ethanediol, and acetate molecules were included in the model. In the last refinement stage, anisotropic B factors were refined to yield an R factor and an $R_{free}$ value of 12.8% and 15.1%, respectively (Table 1). Visual analysis and model manipulation were carried out with O [44]. The relative occupancy of the two Cys65 alternate conformations observed in native crystals was determined by performing separate refinement runs on the protein model, in which the occupancy of each of the two conformations was changed in 10% steps. The fraction that gave the best R factor (i.e., 80% versus 20%) was selected.

The GlpE structures following KCN, $K_2HPO_4$, and $H_3PO_2$ soakings were analyzed based on ($2F_o$-$F_c$) Sim-weighted electron density maps. Phases were calculated from the model structure of native GlpE, in which Cys65 was trimmed to Ala to avoid model bias. Each model was subsequently refined with REFMAC. Electrostatic fields were calculated using DELPHI [19] under the following conditions: pH = 7.5, ionic strength = 0.15 M, $\epsilon_{protein} = 3$, $\epsilon_{solvent} = 80$. The quality of the final model was assessed with PROCHECK [45]. The Ramachandran plot showed that residues are either in the most favored region (91.6%) or in the additionally allowed region (8.4%).

### Structure Comparison and Sequence Analysis

Optimal Cα superposition of protein structures was carried out using programs of Rossman and Argos [46], Diederichs [47], and Kabsch [48]. To identify the *E. coli* ORFs coding for proteins containing rhodanese modules, a PSI-BLAST search was carried out on the sequence database [49]. As a result, 446 ORFs coding for proteins containing at least one rhodanese module, with the putative catalytic Cys at the expected position (including members of the Cdc25 phosphatase family, which were identified by the presence of a longer catalytic loop), were found. Eight of these were part of the *E. coli* K-12 genome [18].

### Acknowledgments

This research was supported in part by Ministero Univesita', Ricerca Scientifica e Tecnologica grant "Solfotransferasi procariotiche", by Agenzia Spaziale Italiana grant IR/167/01, and by an ASPIRES (A Support Program for Innovative Research Strategies) award from Virginia Tech. Data collection at EMBL-Hamburg and ESRF-Grenoble was supported by EU funding.

Received: July 31, 2001
Accepted: September 10, 2001

### References

1. Westley, J., Adler, H., Westley, L., and Nishida, C. (1983). The sulfurtransferases. Fundam. Appl. Toxicol. *3*, 377–382.
2. Nagahara, N., Okazaki, T., and Nishino, T. (1995). Cytosolic mercaptopyruvate sulfurtransferase is evolutionarily related to mitochondrial rhodanese. J. Biol. Chem. *270*, 16230–16235.
3. Sorbo, B. (1955). On the catalytic effect of blood serum on the reaction between colloidal sulfur and cyanide. Acta Chem. Scand. *9*, 5735–5744.
4. Westley, J. (1981). Thiosulfate:cyanide sulfurtransferase (rhodanese). Methods Enzymol. *77*, 285–291.
5. Bonomi, F., Pagani, S., and Kurtz, D.M.J. (1985). Enzymic synthesis of the 4Fe-4S clusters of *Clostridium pasteurianum* ferredoxin. Eur. J. Biochem. *148*, 67–73.
6. Pagani, S., Bonomi, F., and Cerletti, P. (1984). Enzymic synthesis of the iron-sulfur cluster of spinach ferredoxin. Eur. J. Biochem. *142*, 361–366.
7. Nandi, D.L., and Westley, J. (1998). Reduced thioredoxin as sulfur-acceptor substrate for rhodanese. Int. J. Biochem. Cell Biol. *30*, 973–977.
8. Ray, W.K., Zeng, G., Potters, M.B., Mansuri, A.M., and Larson, T.J. (2000). Characterization of a 12-kilodalton rhodanese encoded by *glpE* of *Escherichia coli* and its interaction with thioredoxin. J. Bacteriol. *182*, 2277–2284.
9. Ploegman, J.H., et al., and Russel, J. (1978). The covalent and tertiary structure of bovine liver rhodanese. Nature *273*, 124–129.
10. Bordo, D., Deriu, D., Colnaghi, R., Carpen, A., Pagani, S., and Bolognesi, M. (2000). The crystal structure of a sulfurtransferase from *Azotobacter vinelandii* highlights the evolutionary relationship between the rhodanese and phosphatase enzyme families. J. Mol. Biol. *298*, 691–704.
11. Pagani, S., Forlani, F., Carpen, A., Bordo, D., and Colnaghi, R. (2000). Mutagenic analysis of Thr-232 in rhodanese from *Azotobacter vinelandii* highlighted the differences of this prokaryotic enzyme from the known sulfurtransferases. FEBS Lett. *472*, 307–311.
12. Fauman, E.B., et al., and Saper, M.A. (1998). Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. Cell *93*, 617–625.
13. Reynolds, R.A., et al., and Watenpaugh, K.D. (1999). Crystal structure of the catalytic subunit of Cdc25B required for $G_2$/M phase transition of the cell cycle. J. Mol. Biol. *293*, 559–568.
14. Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A genomic perspective on protein families. Science *278*, 631–637.
15. Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA *95*, 5857–5864.
16. Krafft, T., Gross, R., and Kroger, A. (1995). The function of *Wolinella succinogenenes psr* gene in electron transport with polysulfide as terminal electron acceptor. Eur. J. Biochem. *230*, 601–606.
17. Palenchar, P.M., Buck, C.J., Cheng, H., Larson, T.J., and

Mueller, E.G. (2000). Evidence that ThiI, an enzyme shared between thiamin and 4-thiouridine biosynthesis, may be a sulfurtransferase that proceeds through a persulfide intermediate. J. Biol. Chem. *275*, 8283–8286.
18. Blattner, F.R., et al., and Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. Science *277*, 1453–1474.
19. Gilson, M.K., Rashin, A., Fine, R., and Honig, B. (1985). On the calculation of electrostatic interactions in proteins. J. Mol. Biol. *183*, 503–516.
20. Bordo, D., et al., and Pagani, S. (2001). A persulfurated cysteine promotes active site reactivity in *Azotobacter vinelandii* rhodanese. Biol. Chem. *382*, 1245–1252.
21. Schlesinger, P., and Westley, J. (1974). An expanded mechanism for rhodanese catalysis. J. Biol. Chem. *249*, 780–788.
22. Weng, L., Heinrikson. R.L., and Westley, J. (1978). Active site cysteinyl and arginyl residues of rhodanese. A novel formation of disulfide bonds in the active site promoted by phenylglyoxal. J. Biol. Chem. *253*, 8109–8119.
23. Gilson, M.K., and Honig, B. (1988). Energetics of charge-charge interactions in proteins. Proteins *3*, 32–52.
24. Zang, Z.-Y., and Dixon, J.E. (1993). Active site labeling of the *Yersinia* protein tyrosine phosphatase: the determination of the $pK_a$ of the active site and the function of the conserved histidine 402. Biochemistry *32*, 9340–9345.
25. Colnaghi, R., Pagani, S., Kennedy, C., and Drummond, M. (1996). Cloning, sequence analysis and overexpression of the rhodanese gene of *Azotobacter vinelandii*. Eur. J. Biochem. *236*, 240–248.
26. Hatzfeld, Y., and Saito, K. (2000). Evidence for the existence of rhodanese (thiosulfate:cyanide sulfurtransferase) in plants: preliminary characterization of two rhodanese cDNAs from *Arabidopsis thaliana*. FEBS Lett. *470*, 147–150.
27. Klimmek, O., Kreis, V., Klein, C., Simon, J., Wittershagen, A., and Kroger, A. (1998). The function of the periplasmic Sud protein in polysulfide respiration of *Wolinella succinogenes*. Eur. J. Biochem. *253*, 263–269.
28. Nagahara, N., and Nishino, T. (1996). Role of amino acid residues in the active site of rat liver mercaptopyruvate sulfurtransferase. J. Biol. Chem. *271*, 27395–27401.
29. Nakamura, T., Yamaguchi, Y., and Sano, Y. (2000). Plant mercaptopyruvate sulfurtransferases: molecular cloning, subcellular localization and enzymatic activities. Eur. J. Biochem. *267*, 5621–5630.
30. Papenbrock, J., and Schmidt, A. (2000). Characterizaton of two sulfurtransferase isozymes from *Arabidopsis thaliana*. Eur. J. Biochem. *267*, 5571–5579.
31. Colnaghi, R., Cassinelli, G., Drummond, M., Forlani F., and Pagani, S. (2001). Properties of the *Escherichia coli* rhodanese-like protein SseA: contribution of the active-site Ser240 to sulfur donor recognition. FEBS Lett. *500*, 153–156.
32. Bordo, D., Larson, T.J., Donahue, J.L., Spallarossa, A., and Bolognesi, M. (2000). Crystals of GlpE, a 12 kDa sulfurtransferase from *Escherichia coli*, display 1.06 Å resolution diffraction: a preliminary report. Acta Crystallogr. D *56*, 1691–1693.
33. Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491–497.
34. Otwinowski, Z. (1993). Oscillation data reduction program. In Proceedings of the CCP4 Study Weekend: Data Collection and Processing. L. Sawyer, N. Isaacs, and S. Bayley, eds. (Warrington, UK: SERC Daresbury Laboratory), pp. 56–62.
35. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
36. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.
37. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.
38. Brünger, A.T. (1997). Patterson correlation searches and refinement. Methods Enzymol. *276*, 558–580.
39. Furey, W., and Swaminathan, S. (1997). PHASES-95: a program package for processing and analyzing diffraction data from macromolecules. Methods Enzymol. *277*, 590–620.
40. de la Fortelle, E., and Bricogne, G. (1997). Maximum likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelengh anomalous diffraction methods. Methods Enzymol. *276*, 472–494.
41. Abrahams, J.P., and Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial $F_1$ ATPase. Acta Crystallogr. D *52*, 30–42.
42. Perrakis, A., Sixma, T.K., Wilson, K.S., and Lamzin, V.S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. Acta Crystallogr. D *53*, 448–455.
43. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D *53*, 240–255.
44. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1993). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *50*, 157–163.
45. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the strereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.
46. Rossman, M.G., and Argos, P. (1976). Exploring structural homology of proteins. J. Mol. Biol. *105*, 75–95.
47. Diederichs, K. (1995). Structural superposition of proteins with unknown alignment and detection of topological similarity using a six-dimensional search algorithm. Proteins *23*, 187–195.
48. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. Acta Crystallogr. A *32*, 922–923.
49. Altschul, S.F., et al., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Res. *25*, 3389–3402.
50. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
51. Merrit, E.A., and Murphy, M.E.P. (1994). Raster3d, a program for photorealistic molecular graphics. Acta Crystallogr. D *50*, 869–873.
52. Esnousf, B.M. (1997). An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. J. Mol. Graph. *15*, 138.

## Accession Numbers

Atomic coordinates and structure factors for native and persulfide-free (as obtained after KCN soak) GlpE have been deposited within the RCSB Protein Data Bank and assigned accession numbers 1gmx, r1gmxsf, 1gn0, and r1gn0sf, respectively.