Identification of a putative flexible loop in *Arabidopsis* glutathione synthetase

Chang-Lin WANG* and David J. OLIVER†‡

*Department of Molecular Biology and Biochemistry, University of Idaho, Moscow, ID 83843, U.S.A., and †Department of Botany, Bessey Hall, Iowa State University, Ames, IA 50011, U.S.A.

Glutathione synthetase catalyses the ATP-dependent ligation of γ -glutamylcysteine with glycine to form glutathione. Amino acid sequence comparisons between the *Arabidopsis* and the *Escherichia coli* proteins suggested that a region, identified as a small flexible loop that covers the active site of the *E. coli* protein, might be conserved in the eukaryotic protein. Three site-directed mutations in the *Arabidopsis* protein were generated to test this hypothesis. Two mutations within the conserved region (Lys³⁶⁷/

Pro³⁶⁸ → Asn/Ser and Gly³⁷⁴ → Val) inactivated the enzyme in an *in vivo* assay based on cadmium resistance in *S. pombe*, and in an *in vitro* assay of the activity of the enzyme expressed in *E. coli*. A third mutation outside of this conserved region (Leu³⁶³ → Glu) had a smaller effect in both assays. These results are consistent with the idea that this glycine-rich loop in the *Arabidopsis* and *E. coli* proteins might serve the same function in covering the active site of the enzyme.

INTRODUCTION

Glutathione (GSH) is proposed to serve a central role in protecting plants from environmental stresses [1]. These include, oxidative stress caused by atmospheric air pollutants or excess photosynthetic capacity [2], the toxic effects of xenobiotic organic chemicals [3], and the poisonous effects of some heavy metals [4,5]. A level of resistance to cadmium toxicity is provided by GSH-derived polymers called phytochelatins [6].

GSH synthetase (EC 6.3.2.3) catalyses the ATP-dependent formation of a peptide-bond between the α -carboxyl group of cysteine in γ -glutamylcysteine and the α -amino group of glycine to form GSH. The enzyme belongs to a class of peptide synthetases with acylphosphate intermediates that, in this case, results from the transfer of the γ -phosphate of ATP to the cysteinyl carboxyl group. Attack of this acylphosphate by the α -amino of glycine results in a tetrahedral intermediate, followed by the release of phosphate and the formation of a peptide bond [7]. Other enzymes with similar reaction mechanisms include γ -glutamylcysteine synthetase and D-alanine:D-alanine ligase [8].

The crystal structures of both GSH synthetase [7] and D-alanine:D-alanine ligase [8] from *Escherichia coli* have been solved. While these two enzymes show few similarities in their primary sequence, there is substantial similarity in their three-dimensional structures. In both enzymes, substrate binding occurs in a cleft between the central and C-terminal domains. Substrate binding within this site results in a conformational change in the enzyme, as detected by decreased protease sensitivity. In the D-alanine:D-alanine ligase this structural change involves two flexible loops that fold inward over the large catalytic cavity [8]. These flexible loops are presumed to function by sealing off the reactive acylphosphate intermediate from water, thereby preventing its hydrolysis. The equivalent loops were unresolved in the crystal structure of *E. coli* GSH synthetase [7]. Tanaka et al. [9] have demonstrated that the protease-sensitive site in the *E. coli* GSH synthetase is located within the

larger of these two loops, supporting their central role in the conformational change that follows substrate binding. These authors concluded that lack of resolution in the crystal structure and the changes in protease-sensitivity demonstrate that the loops are highly mobile. Mutagenesis studies have been done on the large flexible loop [9,10,11] and support the essential role of this structure in enzyme activity.

E. coli GSH synthetase is a tetramer of four identical 316 amino acid residue (35.6 kDa) subunits. The enzyme from higher eukaryotes is a homodimer of 53 kDa subunits, composed of 474 amino acid residues in *Xenopus* [12], rat [13] and humans [14], and 478 residues in *Arabidopsis* [15,16]. The protein from *Schizosaccharomyces pombe* has been reported to be a heterodimer of 26 and 33 kDa subunits [17], although recent evidence suggests that there may be problems with this interpretation [16].

Unfortunately, the lack of primary sequence similarity between the prokaryotic and eukaryotic forms of the protein have made it difficult to glean structural information on the plant and animal enzyme from the crystal structure of the *E. coli* enzyme [13]. In this manuscript we report a sequence of the *Arabidopsis* GSH synthetase protein that shares homology with the small flexible loop of the *E. coli* protein, and the results of site-directed mutagenesis experiments that show that this region is essential to catalytic activity.

MATERIALS AND METHODS

Strains and media

The *E. coli* strains used were DH5 α , JM101 and BL21. The *S. pombe* strains used were American Type Culture Collection (ATCC) 38399 and MN101, a mutant selected by Mutoh and Hayashi [18] as being cadmium hypersensitive, and shown to lack GSH synthetase activity. The medium used for *E. coli* was Luria broth. Protein expression in the FLAG-1 vector under control of the *tac* promoter was induced with 0.5 mM isopropyl β -D-thiogalactoside at 30 °C for 2 h. Ampicillin (100 μ g/ml) was

supplied for selection of the plasmid. The medium used for *S. pombe* was yeast extract medium [19]. The *S. pombe* expression vector was pREP3XN [16] which expresses a cloned gene under the control of the *nmt1* promoter and contains the *leu2* gene [8,20,21].

Site-directed mutagenesis of *Arabidopsis* GSH synthetase and complementation of *S. pombe* mutant MN101

Plasmid pGSH222, an *Arabidopsis gsh2* cDNA cloned into pREP3XN, was used for mutagenesis. The primers used for generating mutations were pTCTGAGGCTTCATAACAAAC-AATCC, GAGAAGGCGGAGTTAACAACATCTATGG, pTCCGGGTTTTTCGATAGC, TTGTTTGTTATGAATTC-TCAGAGAGAAG and pGAATTCGTTATGAAGCCTCAG-AGAGAAG. The sequences underlined show the new restriction sites created (*Eco*RI and *Hpa*I). The mutations were generated using the Exsite™ PCR-based site-directed mutagenesis kit (Stratagene, La Jolla, CA, U.S.A.). After the mutated sequences were confirmed by sequencing, the plasmids were introduced into the *S. pombe* mutant, MN101, by electroporation. The transformants were selected for leucine auxotrophy and the presence of the *gsh2* gene was confirmed by colony PCR [16].

Cloning and expression of *Arabidopsis* GSH synthetase in *E. coli*

The FLAG-1 expression system (Scientific Imaging Systems, New Haven, CT, U.S.A.) was used to express the *Arabidopsis* GSH synthetase protein in *E. coli*. The open reading frame of the *Arabidopsis* cDNA clone (wild type and the three mutants) was subcloned from pGSH222 and its derivatives by PCR using two primers, GCGAATTCATGGAATCACAGAAACC and GTTGTCTGACTGATCAAGATGGTTGTGAA (the underlined sequences introduced *Eco*RI and *Sal*I restriction sites). The DNA was then cloned following *Eco*RI (partial)–*Sal*I digestion into the *Eco*RI and *Sal*I sites of the pFLAG-1 plasmid to generate in-frame fusion proteins with the *ompA* export sequence and the FLAG epitope at the N-terminal of the protein.

The fusion proteins were expressed in *E. coli*. The culture containing the recombinant plasmid was grown in 50 ml of Luria broth inoculated with 0.5 ml of an overnight seed culture, and shaken at 37 ° C for 2 h. For standard conditions, 0.5 mM isopropyl β -D-thiogalactoside was added and the culture was shifted to 30 ° C and shaken for an additional 2 h. To extract periplasmic proteins, the bacteria were collected by centrifugation, suspended at 1 g of cells per 40 ml of 0.5 M sucrose/30 mM Tris/HCl (pH 8.0)/1 mM EDTA and kept on ice for 5 min. After centrifugation, the bacteria were resuspended at 1 g per 25 ml of cold water. The supernatant following centrifugation was collected and concentrated 10-fold by ultrafiltration before use.

Enzyme assays

GSH synthetase activity was measured by the ATP-dependent reaction of [^{14} C]glycine with γ -glutamylcysteine to form [^{14} C]GSH, as described previously [16,18]. Enzymic activities are expressed as nmol of ^{14} C-glycine incorporated/h per mg of protein.

RESULTS

Evidence for the putative flexible loop of a substrate-binding site in *Arabidopsis* GSH synthetase

While GSH synthetase from *Arabidopsis* shows substantial sequence similarity with the enzyme from other eukaryotes, the

Table 1 Alignments of eukaryotic and *E. coli* GSH synthetase sequences

Organism	Amino acid sequence
<i>E. coli</i>	154- H S D I I L K P L D G M G G A S I F R
<i>Arabidopsis</i>	361- P G L F V M K P Q R E G G G N N I Y G
Human	358- P S R F V L K P Q R E G G G N N L Y G
Rat	358- P S H F V L K P Q R E G G G N N F Y G
<i>Xenopus</i>	358- P D Q F V L K P Q R E G G G N N L Y G
<i>S. pombe</i>	168- P E D F V L K P Q R E G G G N N T Y G

Table 2 Mutations generated in *Arabidopsis thaliana* GSH synthetase and the plasmids expressing those mutants

The pGSH series of plasmids are all based on pREP3XN and express the protein in *S. pombe*. The pFLAG series are based on pFLAG-1 and are for expression of fusion protein in *E. coli*. The amino acid numbering system is based on the *Arabidopsis* sequence presented in Table 1.

Plasmid	Description
pREP3XN/pFLAG-1	Control plasmids lacking inserts
pGSH222/pFLAG222	Wild-type GSH synthetase
pGSH253/pFLAG253	Leu ³⁶³ → Glu ³⁶³
pGSH259/pFLAG259	Lys ³⁶⁷ → Asn ³⁶⁷
	Pro ³⁶⁸ → Ser ³⁶⁸
pGSH246	Gly ³⁷⁴ → Val ³⁷⁴

similarity with the *E. coli* enzyme has been difficult to establish. Alignment of the known eukaryotic sequences with the *E. coli* protein, using the Macaw computer program [22], suggested that a small region of the *E. coli* amino acid sequence was related to that for the known eukaryote sequences (Table 1). The 11 amino acid *Arabidopsis* GSH synthetase sequence between Phe³⁶⁴ and Gly³⁷⁴ contains three similar and four identical amino acids. While this similarity is not exceptional and might not normally warrant consideration, the similarity overlaps two important regions in the *E. coli* enzyme. Lys³⁶⁷ in the *Arabidopsis* sequence is equivalent to Lys¹⁶⁰ of the *E. coli* sequence, and this residue has been implicated in binding the adenine ring and α -phosphate of ATP [8]. The *Arabidopsis* sequence from Glu³⁷¹ to Gly³⁷⁴ aligns with the *E. coli* sequence from Gly¹⁶⁴ to Gly¹⁶⁷, the smaller flexible loop that has been suggested to fold over the substrate binding site and protect the γ -glutamylcysteinylphosphate intermediate from reaction with bulk water [7,8].

In order to establish that this region in the *Arabidopsis* GSH synthetase is homologous with the proposed region in the *E. coli* sequence, we created a number of site-directed mutations in and around the region, and expressed these modified genes in *S. pombe*. The mutants were generated at three positions, Leu³⁶³ was changed to Glu, Lys³⁶⁷ and Pro³⁶⁸ were simultaneously changed to Asn and Ser, and Gly³⁷⁴ was changed to Val. These mutations are described in Table 2. The mutation in plasmid pGSH253 is outside of the proposed homologous region, whereas the mutations in pGSH246 and pGSH259 are both within the homologous region.

We have previously shown that the *Arabidopsis* cDNA for GSH synthetase can functionally complement the *S. pombe* mutant MN101 that lacks this enzyme activity. Fission yeasts that are unable to synthesize GSH are unable to synthesize phytochelatins. Since phytochelatins are the major source of cadmium resistance in fission yeast (as well as plants), mutations that block

Figure 1 Cadmium resistance of *S. pombe* GSH synthetase-deficient mutant MN101 transformed with different plasmids

1, Wild-type yeast ATCC 38399 transformed with pREP3XN; 2, MN101/pGSH222; 3, MN101/pGSH253; 4, MN101/pGSH259; 5, MN101/pGSH246; 6, MN101/pREP3XN. Panel (A) contains YE medium with no cadmium and (B) contains YE medium with 0.02 mM cadmium.

Figure 2 Western blot of the expression of the *Arabidopsis* GSH synthetase clone pFLAG222 and its derivatives in *E. coli* DH5α

(A) Lanes 1, 3 and 5 are pFLAG222 and lanes 2 and 4 are pFLAG-1 controls. Lanes 1 and 2 are concentrated medium fractions; lanes 3 and 4 are periplasmic-space fractions; and lane 5 is a whole cell extract. (B) The proteins are periplasmic-space preparations from pFLAG222 (lane 1), pFLAG-1 (lane 2), pFLAG246 (lane 3), pFLAG253 (lane 4) and pFLAG259 (lane 5). These are the same fractions assayed in Table 3. The proteins were detected with the anti-FLAG antibody.

phytochelatin biosynthesis result in cadmium hypersensitivity [4,5,18]. Wild-type *S. pombe* grow normally in the presence of 0.5 to 1 mM CdCl_2 . The MN101 mutant lacks endogenous GSH synthetase activity and is sensitive to CdCl_2 concentrations as low as 0.005 mM. Expression of the *Arabidopsis* GSH synthetase

Table 3 Enzymic activity of *Arabidopsis* GSH synthetase and site-directed mutants expressed in *E. coli*

The results expressed are the means $\pm$ S.E.M. of three independent determinations. The enzyme activities were measured either before or after storage at -20°C . A heat-killed control of each preparation showed no activity. The values presented are corrected for the amount of radioactivity found with a periplasmic preparation of *E. coli* containing the pFLAG-1 control plasmid with no insert. The *E. coli* strain used was DH5α, and 20 to 40 μg of protein was used per assay. The specific mutations in each plasmid are shown in Table 2.

Plasmid	Glutathione synthetase activity (nmol/h per mg of protein)	
	Fresh preparation	Frozen preparation
pFLAG-1	0	0
pFLAG222	74.6 ± 1.2	71.9 ± 2.6
pFLAG253	16.2 ± 0.4	16.5 ± 1.0
pFLAG259	0	0
pFLAG246	0	0

cDNA (clone pGSH222) in MN101 restores normal levels of cadmium resistance [16].

In order to evaluate the effects of these mutations on the GSH synthetase activity, the mutated *Arabidopsis gsh2* cDNA clones in the plasmid pREP3XN were transferred into *S. pombe* MN101, and the resulting yeast strains were checked for restoration of cadmium resistance. As shown in Figure 1, in the presence of 0.02 mM cadmium the two mutations located within the homologous region (pGSH246 and pGSH259) are both unable to complement the GSH synthetase deficiency of *S. pombe* MN101. The clone pGSH253, however, which contains a mutation outside of the homologous region, did support some level of growth on cadmium. Based on these preliminary results, Lys³⁶⁷/Pro³⁶⁸ along with Gly³⁷⁴ are essential for function of the GSH synthetase from *Arabidopsis*, while Leu³⁶³ is not as important.

Expression of *Arabidopsis* GSH synthetase in *E. coli*

The level of expression in *S. pombe* was not sufficiently high to allow us to identify the protein on SDS/PAGE gels. This problem was addressed by expressing the wild-type and mutated proteins in *E. coli*. We used the FLAG-1 expression system because it allowed us to use the FLAG epitope to follow expression of the protein. Figure 2 shows a Western blot of *Arabidopsis* GSH synthetase expressed in *E. coli* DH5α. The protein produced was rapidly degraded in the cytosol of *E. coli* (Figure 2A, lane 5). The largest band present, about 54 kDa, is probably the full-length GSH synthetase. The protein contains an N-terminal *ompA* sequence that causes the protein to be exported from the bacteria. The *ompA* sequence is removed during the export process. The periplasmic preparation from these bacteria contained just the 54 kDa protein (Figure 2A, lane 3). Surprisingly, this enzyme from the periplasmic space appeared to be reasonably stable (Table 3). The enzyme from the periplasmic preparation eluted from a Superose 12 column with an apparent molecular mass of 110 kDa, which suggests that it is a homodimer. The same structure has been suggested for the other eukaryotic proteins [13,23].

The degradation of the transgenic GSH synthetase presented a problem because it precluded our obtaining sufficient amounts

of the enzyme to purify it. Several attempts were made to decrease the amount of degradation. These included, lowering induction temperature (30 °C), varying isopropyl β -D-thiogalactoside concentrations for induction (0.1 mM–1.0 mM), altering induction times (0–24 h), supplementing the medium with 0.4 % (w/v) glucose, and replacing the host strain with JM101 and BL21. None of these had any significant positive effect on preventing GSH synthetase degradation. The protein was also expressed in the maltose expression system (Biolab, MA, U.S.A.) without improvement. Although we were unable to stabilize the protein to allow purification, we were able to isolate sufficient protein from the periplasmic preparations to measure the activity of the enzyme.

The results from the GSH synthetase activity measurements are shown in Table 3. The controls were from DH5 α containing the control pFLAG-1 plasmid with no insert. The enzyme from tobacco had a K_m for glycine of 0.4 mM [23], while the wild-type transgenic *Arabidopsis* protein had a K_m of 0.8 mM.

The *Arabidopsis* cDNA for GSH synthetase produced a substantial amount of enzyme activity. The plasmid pFLAG253 containing the *Arabidopsis* GSH synthetase with the Leu²⁶³ $\rightarrow$ Glu mutation showed about 22 % of the wild-type activity. Both mutations within the homologous region, pFLAG 259 (Lys³⁶⁷/Pro³⁶⁸ $\rightarrow$ Asn/Ser) and pFLAG246 (Gly³⁷⁴ $\rightarrow$ Val), produced proteins that had no detectable enzyme activity. Using the anti-FLAG antibody, it was possible to show that nearly equal amounts of protein were present in each reaction (Figure 2B). The results for the protein isolated from *E. coli*, therefore, agree with those measuring *in situ* levels of enzyme activity in *S. pombe*, as indicated by cadmium resistance. The mutations within the region of apparent similarity result in complete inhibition of the *Arabidopsis* protein.

DISCUSSION

The lack of apparent sequence homology between the GSH synthetase protein from *E. coli* and that from eukaryotes has made it difficult to exploit the structural information gained on the prokaryotic protein to understand the protein from higher organisms. The region of limited sequence similarity studied is highly conserved in the eukaryotic proteins, suggesting that it contains a sequence that is important for enzymic activity. Our site-directed mutagenesis study confirmed that this region is important. Both mutations within the putative flexible loop region (Lys³⁶⁷/Pro³⁶⁸ $\rightarrow$ Asn/Ser and Gly³⁷⁴ $\rightarrow$ Val) resulted in the production of proteins that were inactive *in vivo* (in the *S. pombe* cadmium-resistance assay) and *in vitro* (following expression in *E. coli*). A mutation outside of this conserved region, Leu³⁶³ $\rightarrow$ Glu, decreased activity only 78 % and resulted in an enzyme that was still active in both the *in vivo* and *in vitro* assays.

This glycine-rich loop in *Arabidopsis* protein could contain the psi and phi angle flexibility between the amino acids to allow movement in the putative flexible loop. The substitution of a

bulky valine residue for Gly³⁷⁴ would be expected to inhibit rotational freedom in the region, and should cause loss of function in a sequence where conformational flexibility is essential. Similar substitutions of valine in the larger flexible loop of the *E. coli* GSH synthetase also resulted in loss of enzyme activity [9,11]. The ability to align these sequences in the *Arabidopsis* and *E. coli* amino acid sequences for GSH synthetase could provide a basis for identifying other potentially important regions in the sequence of this important protein from eukaryotes.

The *Arabidopsis* GSH synthetase protein seems amenable to further study. It is active when expressed in *E. coli*. Rawlins et al. [15] have shown that it is capable of functioning in *E. coli in vivo* and could complement a mutation in the bacterial *gsh2* locus. We showed earlier [16] that the enzyme functioned in *S. pombe*. In the present study we have demonstrated that the protein exported from *E. coli* has chromatographic, electrophoretic and kinetic properties that are very similar to the enzyme isolated from plants.

REFERENCES

- 1 Alscher, R. G. (1989) *Physiol. Plant.* **77**, 457–464
- 2 Bergmann, L. and Rennenberg, H. (1993) In *Sulfur Nutrition and Assimilation in Higher Plants* (De Kok, L. D., Stulen, I., Rennenberg, H., Brunold, C. and Rauser, W. E., eds.), pp. 109–123. SPB Academic Publishing bv, The Hague
- 3 Lamoureux, G. L., Shimabukuro, R. H. and Frear, D. S. (1994) In *Herbicide Resistance in Weeds and Crops* (Caseley, J. C., Cussans, G. W. and Atkin, R. K., eds.), pp. 227–261. Butterworth Heinemann, Oxford
- 4 Howden, R., Andersen, C. R., Goldsbrough, P. B. and Cobbett, C. S. (1995) *Plant Physiol.* **107**, 1067–1073
- 5 Howden, R., Goldsbrough, P. B., Andersen, C. R. and Cobbett, C. S. (1995) *Plant Physiol.* **107**, 1059–1066
- 6 Rauser, W. E. (1995) *Plant Physiol.* **109**, 1141–1149
- 7 Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J. and Katsube, Y. (1993) *Mol. Biol.* **229**, 1083–1100
- 8 Fan, C., Noews, P. C., Shi, Y., Walsh, C. T. and Knox, J. R. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1172–1176
- 9 Tanaka, T., Yamaguchi, H., Kato, H., Nishioka, T., Katsube, Y. and Oda, J. (1993) *Biochemistry* **32**, 12398–12404
- 10 Kato, H., Tanaka, T., Yamaguchi, H., Hara, T., Nishioka, T., Katsube, Y. and Oda, J. (1994) *Biochemistry* **33**, 4995–4999
- 11 Tanaka, T., Kato, H., Nishioka, T. and Oda, J. (1992) *Biochemistry* **31**, 2259–2265
- 12 Habenicht, A., Hille, S. and Knochel, W. (1993) *Biochim. Biophys. Acta* **1174**, 295–298
- 13 Huang, C.-S., He, W., Meister, A. and Anderson, M. E. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1232–1236
- 14 Gali, R. R. and Board, P. G. (1995) *Biochem. J.* **310**, 353–358
- 15 Rawlins, M. R., Leaver, C. J. and May, M. J. (1995) *FEBS Lett.* **376**, 81–86
- 16 Wang, C. and Oliver, D. J. (1996) *Plant. Mol. Biol.* **31**, 1093–1104
- 17 Mutoh, N., Nakagawa, C. W., Ando, S., Tanabe, K. and Hayashi, Y. (1991) *Biochem. Biophys. Res. Commun.* **181**, 430–436
- 18 Mutoh, M. and Hayashi, Y. (1988) *Biochem. Biophys. Res. Commun.* **151**, 32–39
- 19 Alfa, C., Fantes, P., Hyams, J., McLeod, M. and Warbrick, E. (1993) *Experiments with Fission Yeast*, Cold Spring Harbor Laboratory Press, New York
- 20 Basl, G., Schmid, E. and Maundrell, K. (1993) *Gene* **123**, 131–136
- 21 Maundrell, K. (1993) *Gene* **123**, 127–130
- 22 Schuler, G. D., Altschul, S. F. and Lipman, D. J. (1991) *Proteins: Struct. Funct. Genet.* **9**, 180–190
- 23 Hell, R. and Bergmann, L. (1988) *Physiol. Plant.* **72**, 70–77