FEBS 16324 FEBS Letters 376 (1995) 81–86

Characterisation of an *Arabidopsis thaliana* cDNA encoding glutathione synthetase

Marion R. Rawlins[a], Christopher J. Leaver[a], Mike J. May[b,*]

[a]*Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, UK, OX1 3RB*
[b]*Laboratorium voor Genetica, Universiteit Gent, K.L. Ledeganckstraat 35, B-9000, Gent, Belgium*

Received 19 July 1995; revised version received 19 October 1995

Abstract An *Arabidopsis thaliana* cDNA (AtGSHS) encoding a protein with high primary sequence identity to cDNAs previously isolated from *Xenopus laevis* (42%), *Schizosaccharomyces pombe* (40%), *Rattus norvegicus* (40%) and *Homo sapiens* (37%) encoding glutathione synthetase (EC 6.3.2.3) has been isolated by functional complementation of an *Escherichia coli* mutant deficient in this enzyme. AtGSHS is encoded by a single gene, *GSHB*, as determined by Southern blot analysis and the corresponding mRNA is abundant in both roots and leaves of *Arabidopsis*.

Key words: *Arabidopsis thaliana*; Glutathione synthetase; Glutathione; Functional complementation

1. Introduction

Plants must respond to adverse environmental conditions through increases in the activity of mechanisms which maintain cellular homeostasis. A result of both biotic and abiotic stresses is often an increased production of partially reduced oxygen species. Accumulation of these highly reactive oxygen intermediates (ROI's) lead to the state of oxidative stress during which considerable damage may be inflicted upon the plant at the cellular level [1]. Cellular protection under these conditions is afforded in part through the activity of antioxidant enzymes (superoxide dismutases, catalases and peroxidases) and antioxidant molecules (ascorbate, carotenoids, tocopherols and glutathione) which together act to limit the accumulation of ROIs and thus reduce oxidative damage [2]. Plant antioxidant defences have received much attention because of their central role in metabolism during conditions of environmental stresses [3]. However, although increases in the activity of antioxidant enzymes and the levels of antioxidant molecules can be measured during early responses to oxidative stress, these changes are often not sufficient to fully protect macromolecules sensitive to oxidative damage. Considerable effort has thus been directed to the genetic manipulation of plant antioxidant defences in order to produce greater tolerance to oxidative stress. Attempts to date have been largely concerned with the manipulation of specific enzymatic antioxidants, in particular superoxide dismutases, ascorbate peroxidase and glutathione reductase [4]. The level of tolerance to oxidative stress which has been achieved has, however, been variable and there is a clear need to gain a better understanding of the cellular antioxidant defences as a whole. The tripeptide thiol glutathione (GSH) is intimately involved in the redox balance of the cell and in the regeneration of ascorbate from dehydroascorbate [5]. As such it is an integral part of the cellular antioxidant system and its biosynthesis occurs in response to oxidative stimuli in plants and during the adaptation of plants to environmental adversity such as drought and extremes of temperature [6,7]. GSH has been identified with a number of protective roles within the plant. It functions in the storage and transport of sulphur, and in the detoxification of xenobiotics and removal of heavy metals. It is associated with protein stability, and exerts redox control over enzyme activity. Given the importance of these diverse yet interrelated roles, tight regulation of GSH level within the cell must be vital to normal cell function in conferring the capability for increased redox buffering capacity during oxidative stress. Despite this, only limited biochemical information is available concerning the structural properties or the regulation of the enzymes responsible for its synthesis. The processes controlling GSH metabolism thus present an attractive target for investigation and manipulation.

GSH is synthesised in two ATP-dependent steps. In the first, catalysed by γ-glutamylcysteine synthetase (γ-ECS; EC 6.3.2.2) the dipeptide γ-glutamylcysteine is produced from L-glutamic acid and L-cysteine. In the second step, catalysed by glutathione synthetase (GSHS; EC 6.3.2.3), glycine is added to the C-terminal site of γ-glutamylcysteine to form GSH. We have adopted a cDNA cloning approach to understand the pathway responsible for the synthesis of GSH in *Arabidopsis*. This will provide the tools with which to investigate the factors which regulate the expression of this pathway and for the genetic manipulation of GSH homeostasis in plants. To this end we have previously described the isolation and characterisation of a cDNA encoding chloroplastic γ-glutamylcysteine synthetase [8]. In this paper, we report the isolation of a cDNA clone encoding glutathione synthetase obtained by functional complementation of an *E. coli* mutant deficient in this enzyme [9].

2. Materials and methods

2.1. Plant material

Arabidopsis thaliana accession Columbia C24 was grown in soil under 16-h light conditions at 21°C. Plants for total RNA isolation were grown as previously described [8].

2.2. Bacterial strains and cDNA library

E. coli K12 parental strain AB1157 (F^-, *thr, leu, proA, his, argB, thi, strA*), and GSHS deficient *gshB* mutant 830 were a kind gift of

*Corresponding author. Fax (32) (9) 2645349.
BITNET: mimay@gengenp.rug.ac.be

The nucleotide sequence reported in this paper has been deposited in the EMBL database with accession number Z50153.

Abbreviations: AtGSHS, *Arabidopsis thaliana* glutathione (*GSH*) synthetase.

0014-5793/95/$9.50 © 1995 Federation of European Biochemical Societies. All rights reserved.
SSDI 0014-5793(95)01253-2

P. Apontoweil [9]. The *Arabidopsis thaliana* cDNA plasmid expression library in λYES-R was a gift from R. Davis [10].

2.3. Isolation and characterisation of cDNA clones

Preparation of *gsh*B mutant 830 for electroporation and electrotransformation was as described previously [8]. After electroporation using 1 μg of the cDNA library, cells were plated on minimal medium containing 0.8% (w/v) agar and 1 × M9 salts [11, appendix A3] supplemented with 8-hydroxyquinoline (8-HQ) at 20 $\mu g \cdot ml^{-1}$, ampicillin at 100 $\mu g \cdot ml^{-1}$, 1 mM isopropyl B-D-thiogalactopyranoside (IPTG), and glucose at 0.1% (w/v). Colonies which appeared after 4 days of incubation at 37°C were selected, grown in L-broth and their glutathione content determined as described [7]. One clone, M12, in which the extractable glutathione content was significantly higher than that measured in extracts of *gsh*B mutant 830 was selected for further analysis. The sequence of the complementing plasmid insert was determined on both strands by subcloning the cDNA insert from pYES-R into pBluescript II KS(+) (Stratagene) and using the dideoxynucleotide chain termination method [12]. Sequence analysis was performed by using the University of Wisconsin Genetics Computer Group software package. The cDNA was named AtGSHS. The pYES plasmid isolated containing AtGSHS was named pM12A. AtGSHS subcloned into pKS(+) was named pM12B.

2.4. DNA and RNA blot hybridisation analysis

Arabidopsis DNA was extracted as described by Dellaporta et al. [13], and total RNA was isolated as described by Gurr and McPherson [14]. Separation of DNA and RNA using agarose gel electrophoresis was by standard procedures [11]. For both analyses, 1 kb and 0.6 kb fragments of clone pM12B produced by digestion with *Xho*I were used together as the probe. Prehybridisation and hybridisation of the filters after transfer was as described [8] followed by low stringency washing. The genomic Southern blot was then washed at high stringency.

2.5. pcr amplification of 5′ sequence

Amplification of DNA by nested pcr was carried out on the p λYES library essentially as described [15]. Sequence specific primers were used in conjunction with primers designed to either the *lac* or *GAL 1* promoters present in p λYES [10].

3. Results

3.1. Isolation of an Arabidopsis *GSHS cDNA by functional complementation*

Electrotransformation of an *Arabidopsis* cDNA plasmid expression library into *gsh*B mutant 830, an *E. coli* mutant deficient in glutathione synthetase (GSHS) activity, yielded thirty clones with restored resistance to 20 $\mu g \cdot ml^{-1}$ 8-HQ. Following glutathione (GSH) content determination one of them (M12) in which the GSH content was equal to the wild type (AB1157) was taken for further analysis. Curing of the plasmid pM12A from isolate M12 by repeated subculturing on media lacking ampicillin and retransformation of the deficient mutant with pM12A extracted from M12, confirmed the plasmid dependence of both 8-HQ resistance (Fig. 1) and restoration of GSH level (Table 1).

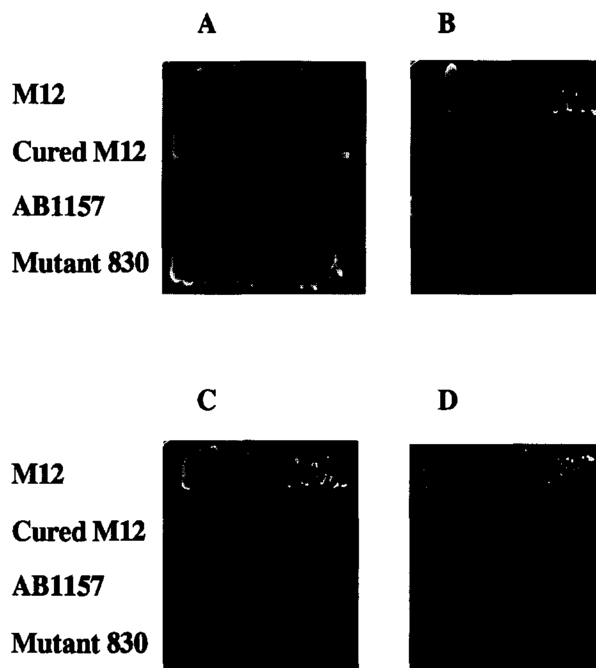

Fig. 1. Complementation of 8-hydroxyquinoline (8-HQ) sensitivity in GSHS-deficient *E. coli* mutant 830 with pM12A. Strains shown are the complemented mutant (M12), cured complemented mutant (cured M12), wild type (AB1157), and GSHS-deficient mutant (mutant 830). (A) Growth on minimal medium (MM) alone. (B) Growth on MM supplemented with 20 $\mu g \cdot ml^{-1}$ 8-HQ. (C) Growth on MM supplemented with 20 $\mu g \cdot ml^{-1}$ 8-HQ and 1 mM IPTG. (D) Growth on MM supplemented with 20 $\mu g \cdot ml^{-1}$ 8-HQ, 1 mM IPTG and 100 $\mu g \cdot ml^{-1}$ ampicillin.

Table 1
Restoration of glutathione levels in a GSHS-deficient *E. coli* mutant by complementation with pM12A

Strain	n mol GSH 10^9 bacteria^{-1}
AB1157	6.21 (± 1.40)
AB1157*	5.85 (± 1.36)
Mutant 830	0.41 (± 0.13)
Mutant 830*	0.41 (± 0.11)
M12	5.42 (± 0.28)
M12*	5.91 (± 0.57)
Cured M12	0.50 (± 0.07)
Cured M12*	0.46 (± 0.16)

Quantification of glutathione (GSH) in *E. coli* wild type (AB1157), GSHS-deficient mutant (mutant 830), mutant complemented with *Arabidopsis* clone AtGSHS in pM12A (M12) and cured complemented mutant (cured M12). Cultures of strains denoted by * were induced with 1 mM IPTG for 2 h before harvesting. Values are the means of three experiments with two replicates per experiment, ± the standard error of the mean.

3.2. Glutathione levels of complemented GSHS-deficient E. coli *mutant*

The glutathione content of the wild type; AB1157, mutant; *gsh*B 830, the mutant complemented with the AtGSHS cDNA; M12, and the complemented mutant from which the plasmid had been cured; cured M12, was determined in extracts made from exponentially growing cultures (Table 1). No significant differences were observed between the glutathione content of the deficient mutant and the cured complemented mutant. However, the mutant containing pM12A grown in the presence or absence of 1 mM IPTG contained levels of glutathione equal to those from the wild type. Taken together, these lines of evidence confirm that the cDNA contained in pM12A is responsible for the restoration of glutathione levels in a mutant which lacks glutathione synthetase activity and contains insignificant levels of glutathione.

3.3. Molecular characterisation of an Arabidopsis *GSHS cDNA*

The sequence of the cDNA insert was determined to be 1671 bp with one open reading frame of 1577 bp, which encodes a

```
     GAGCTACGTCAGGGACTCAATCCAAGTAGCTTCCTTTTCCAGAACCCTAAAACTCTGAGA   60
  1  E  L  R  Q  G  L  N  P  S  S  F  L  F  Q  N  P  K  T  L  R

     AATCAGTCACCTTTGAGGTGTGGGAGATCATTTAAGATGGAATCACAGAAACCCATTTTC  120
 21  N  Q  S  P  L  R  C  G  R  S  F  K  M  E  S  Q  K  P  I  F

     GATTTGGAGAAATTAGATGATGAGTTTGTTCAGAAACTGGTTTACGATGCTCTCGTTTGG  180
 41  D  L  E  K  L  D  D  E  F  V  Q  K  L  V  Y  D  A  L  V  W

     TCTTCTCTTCACGGACTCGTCGTTGGTGACAAAAGTTATCAGAAATCAGGAAATGTTCCA  240
 61  S  S  L  H  G  L  V  V  G  D  K  S  Y  Q  K  S  G  N  V  P

     GGGGTTGGATTGATGCACGCACCTATTGCATTGCTACCAACTGCATTCCCAGAAGCTTAT  300
 81  G  V  G  L  M  H  A  P  I  A  L  L  P  T  A  F  P  E  A  Y

     TGGAAGCAAGCTTGTAATGTTACTCCTCTTTTCAATGAATTGATTGATCGTGTTAGCTTG  360
101  W  K  Q  A  C  N  V  T  P  L  F  N  E  L  I  D  R  V  S  L

     GATGGCAAATTCTTACAGGATAGTCTCTCTAGAACGAAGAAAGTTGATGTCTTTACATCT  420
121  D  G  K  F  L  Q  D  S  L  S  R  T  K  K  V  D  V  F  T  S

     AGACTTCTTGACATTCACTCCAAGATGCTAGAAAGAAATAAGAAAGAGGACATTCGTTTG  480
141  R  L  L  D  I  H  S  K  M  L  E  R  N  K  K  E  D  I  R  L

     GGTTTACACCGGTTTGATTATATGCTTGACGAAGAAACAAATTCACTTCTTCAGATTGAG  540
161  G  L  H  R  F  D  Y  M  L  D  E  E  T  N  S  L  L  Q  I  E

     ATGAATACTATCTCGTGTTCGTTTCCTGGCCTTAGTCGTCTTGTTAGCCAGCTACATCAG  600
181  M  N  T  I  S  C  S  F  P  G  L  S  R  L  V  S  Q  L  H  Q

     TCATTGCTTCGATCTTATGGGGATCAGATTGGCATAGACTCTGAACGTGTACCTATAAAC  660
201  S  L  L  R  S  Y  G  D  Q  I  G  I  D  S  E  R  V  P  I  N

     ACATCCACAATCCAATTTGCTGATGCATTGGCTAAAGCTTGGTTGGAGTACAGTAACCCA  720
221  T  S  T  I  Q  F  A  D  A  L  A  K  A  W  L  E  Y  S  N  P

     AGAGCGGTAGTCATGGTAATTGTACAGCCAGAAGAACGCAACATGTACGATCAACATTTG  780
241  R  A  V  V  M  V  I  V  Q  P  E  E  R  N  M  Y  D  Q  H  L

     CTGAGCAGTATATTGAGAGAAAAGCATAATATAGTTGTCATCAGGAAGACTCTAGCAGAA  840
261  L  S  S  I  L  R  E  K  H  N  I  V  V  I  R  K  T  L  A  E

     GTCGAAAAAGAAGGGAGTGTACAAGAGGATGAAACCCTTATTGTTGGCGGCCAAGCAGTC  900
281  V  E  K  E  G  S  V  Q  E  D  E  T  L  I  V  G  G  Q  A  V

     GCAGTGGTTTATTTCAGATCTGGCTATACTCCTAATGATCATCCGTCTGAATCAGAATGG  960
301  A  V  V  Y  F  R  S  G  Y  T  P  N  D  H  P  S  E  S  E  W

     AATGCTAGGCTGCTTATTGAGGAGTCTTCAGCTGTCAAATGCCCGAGCATAGCTTACCAT 1020
321  N  A  R  L  L  I  E  E  S  S  A  V  K  C  P  S  I  A  Y  H

     TTAACTGGCTCCAAGAAAATCCAGCAAGAACTGGCTAAACCAGGTGTTCTCGAGAGGTTT 1080
341  L  T  G  S  K  K  I  Q  Q  E  L  A  K  P  G  V  L  E  R  F

     CTGGACAACAAAGAGGACATTGCTAAGCTGAGGAAATGCTTTGCTGGGCTTTGGAGCTTG 1140
361  L  D  N  K  E  D  I  A  K  L  R  K  C  F  A  G  L  W  S  L

     GATGACTCAGAAATTGTCAAGCAGGCTATCGAAAAACCCGGATTGTTTGTTATGAAGCCT 1200
381  D  D  S  E  I  V  K  Q  A  I  E  K  P  G  L  F  V  M  K  P

     CAGAGAGAAGGCGGAGGAAACAACATCTATGGAGATGATGTGAGGGAAAATCTTTTGAGA 1260
401  Q  R  E  G  G  G  N  N  I  Y  G  D  D  V  R  E  N  L  L  R

     CTGCAGAAAGAAGGAGAGGAAGGAAACGCTGCGTATATCCTGATGCAGAGGATATTCCCA 1320
421  L  Q  K  E  G  E  E  G  N  A  A  Y  I  L  M  Q  R  I  F  P

     AAAGTCTCAAACATGTTCTTGGTGCGAGAAGGCGTTTACCATAAGCATCAAGCTATATCA 1380
441  K  V  S  N  M  F  L  V  R  E  G  V  Y  H  K  H  Q  A  I  S

     GAACTCGGTGTCTATGGTGCTTACCTCAGGAGCAAAGACGAAGTTATAGTAAACGAGCAG 1440
461  E  L  G  V  Y  G  A  Y  L  R  S  K  D  E  V  I  V  N  E  Q

     AGTGGTTATCTAATGCGCACGAAGATCGCATCATCAGATGAAGGCGGCGTTGCAGCTGGT 1500
481  S  G  Y  L  M  R  T  K  I  A  S  S  D  E  G  G  V  A  A  G

     TTTGGAGTCTTGGACAGCATATATCTGATTTGAgatggataacattcacaaccatcttga 1560
501  F  G  V  L  D  S  I  Y  L  I  *

     tcagtttttcttctggtttattccttcttcttctcttcttgccttgttaaagatttgtgt 1620

     attaatcatgaagatcaataagactaggcaaagcccctgacgtagctcg            1669
```

Fig. 2. Nucleotide and derived amino acid sequences of the cDNA AtGSHS in pM12B. The nucleic acid sequence is presented on the top line with the derived amino acid sequence below. Amino acid residues are numbered on the left; nucleotide positions are numbered on the right. The 3′ untranslated region is shown in lower case type. Upper case type indicates the sequence from the beginning of the clone to the stop codon TGA.

protein of 510 amino acids (Fig. 2) with a calculated molecular mass of 58 kDa.

The derived amino acid sequence of the AtGSHS shared high homology with the derived amino acid sequences for GSHS from several other species of eukaryotes. Computer-aided comparison revealed 42% identity with *Xenopus laevis* [16], 40% with *Rattus norvegicus* [17] and the large subunit of the *Schizosaccharomyces pombe* enzyme [18], and 37%, 100% and 95% with unpublished sequences from *Homo sapiens*, and two from *Arabidopsis*, respectively (Table 2). This high homology extends over the entire amino acid sequence but is most apparent between amino acids 409 and 424 on the AtGSHS sequence, (at 1 in Fig. 3) where all seven sequences are virtually identical (Fig. 3). The AtGSHS shared only 16% identity with the GSHS of *E. coli* [19] and *Anaplasma centrale* [20].

```
     1                                                    50
AT1  ELRQGLNPSS FLFQNPKTLR NQSPLRCGRS FKMESQKPIF DLEKLDDEFV
AT2  .......... .......... ...PLRCGRS FKMESQKPIF DLEKLDDEFV
AT3  .......... .LFQNPKTLR NQPPLRCGRS FKMESQKPIF DLEKLDDEFV
 XL  .......... .......EFH ESELVR.GLH LSRIRMADLW DDIYNDTKLL
 RN  .......... .......... .ECELGEQLD NERVGMATSW GSILQDEKQL
 SP  .......... .......... .......... .......... ..........
 HS  .......... .......... .......... .......... ..........

     51                                                  100
AT1  QKLVYDALVW SSLHGLVVGD KSYQKSGNVP GVGLMHAPIA LLPTAFPEAY
AT2  QKLVYDALVW SSLHGLVVGD KSYQKSGNVP GVGLMHAPIA LLPTAFPEAY
AT3  QKLVYDALVW SSLHGLGVGD KTYQKSGNVP GVGLMHAPIA LLPTGFPEVL
 XL  EELAPIAIDA ALLQGVLMRT KESPNSSDV. ...VSFAPFA LLPSPVPKAL
 RN  EELAQQAIDR ALAEGVLLRS AKNPSSSDV. ...VTYAPFT LFPSPVPSTL
 SP  .......... .......... .......... .......... ..........
 HS  .......... .......... .......... .......... ..........

     101                                                 150
AT1  W.KQACNVTP LFNELIDRVS LDGKFLQDSL SRTKKVDVFT SRLLDIHSKM
AT2  W.KQACNVTP LFNELIDRVS LDGKFLQDSL SRTKKVDVFT SRLLDIHSKM
AT3  TGSQACDVTP LFNELIDRVS LDGKFLQDSL SRTKKVDVFT SRLLDIHSKM
 XL  F.EQAKCVQE DFNTLVDRIS QDTSFLEQVL SSTIKVDDFI RRLFAIHKQV
 RN  L.EQAYAVQM DFNILVDAVS QNSAFLEQTL SSTIKKDEYT ARLFDIYKQV
 SP  .......... .......... .......... .......... ..........
 HS  .......... .......... .......... .......... ..........

     151                                                 200
AT1  LERNKKEDIR LGLHRFDYML D...EETNSL LQIEMNTISC SFPGLSRLVS
AT2  LERNKKEDIR LGLHRFDYML D...EETNSL LQIEMNTISC SFPGLSRLVS
AT3  LERNKKEDIR LGLRRFDYML D...EETNSL LQIEMNTISC SFPALSRLVS
 XL  QQEDCTQEVF LGINRSDYMF DCRDDGTPAL KQIEINTIAA SFGGLASRTP
 RN  LKEGIAQTVF LGLNRSDYMF QCSADGSKAL KQIEINTISA SFGGLASRTP
 SP  .......... .......... .......... .......... ..........
 HS  .......... .......... .......... .......... ..........

     201                                                 250
AT1  QLHQSLLRSY GDQIGIDSER VPINTSTIQF ADALAKAWLE YSNPRAVVMV
AT2  QLHQSLLRSY GDQIGIDSER VPINTSTIQF ADALAKAWLE YSNPRAVVMV
AT3  QLHQSLLRSY GDQIGIDSER VPINTSTIQF ADALAKAWLE YSNPRAVVMV
 XL  AVHQHVLKFL RK..SEESSS ILTNDAVEGI GWGIAHAWAL YGSVDATVMF
 RN  AVHRHVLNVL NK..TNEASK ILSNNPSKGL ALGIAKAWEL YGSANAVVLL
 SP  .......... .......... .......... ......MNIA SDNTKPIVLF
 HS  .......... .......... .......... .......... ..........

     251                                                 300
AT1  IVQPEERNMY DQHLLSSILR EKHNIVVIRK TLAEVEKEGS VQEDETLIVG
AT2  IVQPEERNMY DQHLLSSILR EKHNIVVIRK TLAEVEKEGS VQEDETLIVG
AT3  IVQPEERNMY DQHLLSSILR EKHNIVVIRK TLAEVEKEGS VQEDETLIVG
 XL  LVENEQRNIL DQRFIEAELC KR.NVRVIRR RLADVFERGT LDEERHLFID
 RN  IAQEKERNIP DQRAIENELL DR.KIHVIRR RFEDVSERGS LDQNRRLFME
 SP  VVKGGERNIT DQRTLEYELL NRFHVISKRI DIAELTSLIH DKSSNKLYMK
 HS  .......... .......... .......... .......... ..........

     301                                                 350
AT1  GQ....AVAV VYFRSGYTPN DHPSESEWNA RLLIEESSAV KCPSIAYHLT
AT2  GQ....AVAV VYFRSGYTPN DHPSESEWNA RLLIEESSAV KCPSIAYHLT
AT3  GQ....AVAV VYFRSGYTPN DHPSESEWNA RLLIEESSAV KCPSIAYHLT
 XL  GY....EVAV AYFRTGYVPQ DYTEQ.DWEA RLMLERSRAV KCPDVPTQLV
 RN  DQ....EVAV VYFRDGYMPS QYNAQ.NWEA RLLLERSCAA KCPDIATQLA
 SP  TSFTTYEVAV VYYRVGYALD DYPSQERWDM RLTIENTLAI KCPSISTHLA
 HS  .......... .......... .......... .......... ..........

     351                                                 400
AT1  GSKKIQQELA KPGVLERFLD NK.EDIAKLR KCFAGLWSLD DS....EIVK
AT2  GSKKIQQELA KPGVLERFLD NK.EDIAKLR KCFAGLWSLD DS....EIVK
AT3  GSKKIQQELA KPGVLERFLD NK.EDIAKLR KCFAGLWSLD DS....EIVK
 XL  GTKKVQQELS RPQILEKFLP DKPEAVARIR ETFTGLYSLD IGEEGDEAVR
 RN  GTKKVQQELS RVGLLEALLP GQPEAVARLR ATFAGLYSLD MGEEGDQAVA
 SP  GSEKIQQVLA ESNALERFLE G..DELQAVR STFADMYPLD DTPRGKEGIK
 HS  .......... .......... .......... .........M WVKKGTRPSP

     401                                                 450
AT1  QAIEKPGLFV MKPQREGGGN NIYGDDVREN LLRLQKEGEE GNAAYILMQR
AT2  QAIEKPGLFV MKPQREGGGN NIYGDDVREN LLRLQKEGEE GNAAYILMQR
AT3  QAIEKPGLFV MKPQREGGGN NIYGDDVREN LLRLQKEGEE GNAAYILMQR
 XL  VALANPDQFV LKPQREGGGN NLYGEELKEK LQECK..DSE ERTSYILMDK
 RN  EALAAPSHFV LKPQREGGGN NFYGEEMVHA LEQLK..DSE ERASYILMEK
 SP  LAFEKPEDFV LKPQREGGGN NTYGKDIPGL LSKM...PQE EWDSYILMRY
 HS  MALAAPSRFV LKPQREGGGN NLYGEEMVQA LKQLK..DSE ERASYILMEK

     451                                                 500
AT1  IFPKVSNMFL VR.EGVYHKH QAISELGVYG AYLR..SKDE VIVNEQSGYL
AT2  IFPKVSNMFL VR.EGVYHKH QAISELGVYG AYLR..SKDE VIVNEQSGYL
AT3  IFPKVSNMFL VR.EGVYHKH QAISELGVYG AYLR..SKDK VIVNEQSGYL
 XL  INPKPLKNCL LRAGGRVQIS ECISELGMFG VYVRH..RDQ MIYYDQVGHL
 RN  IEPEPFRNCL LRPGSPAQVV QCISELGIFG VYVRQ..GTT LVMNKHVGHL
 SP  INAVPSQNYI LK.GERPEKF DVVDEIGILG TIVWNINTDE VVQNGQSGFI
 HS  IEPEPFENCL LRPGSPARVV QCISELGIFG VYVRQ..EKT LVMNKHVGHL

     501                                                 550
AT1  MRTKIASSDE GGVAAG.FGV LDSIYLI*.. .......... ..........
AT2  MRTKIASSDE GGVAAG.FGV LDSIYLI*
AT3  MRTKIASSDE GALQLV.LEP WTAFI*
 XL  LRTKAIEHSD GGVAAGV.AV LDNPYLV*.. .......... ..........
 RN  LRTKAIEHAD GGVAAGV.AV LDNPYPV*.. .......... ..........
 SP  CRTKPKKTNE GGVATG.YAS LSSIELSE*. .......... ..........
 HS  LRTKAXEHAD GXFGSGSWQS WDNPYPVFKG TTRXR..... ..........
```

Fig. 3. Comparison of the derived amino acid sequence of AtGSHS (AT1) with that from *X. laevis* (XL) [14], *R. norvegicus* (RN) [15], *S. pombe* (SP) [16], *H. sapiens* (HS), and two other *Arabidopsis* sequences present in the EMBL database (AT2 and AT3). Translations were made of nucleic acid sequences present in the EMBL database. The sequences are aligned for maximal homology, gaps are represented as dots and amino acids identical to those in the *Arabidopsis* sequence are highlighted in bold.

Genomic Southern blot analysis was carried out using *Arabidopsis* DNA digested with *Hind*III, *Xho*I and *Pst*I, which have restriction sites within AtGSHS, or *Eco*RI and *BamH*I, which do not. The DNA fragments detected by the cDNA probe after low stringency washing remained at high stringency. The simple pattern of hybridisation suggests that the gene corresponding to AtGSHS is present as a single copy in the *Arabidopsis* genome (Fig. 4). We have designated this locus *GSHB*. This locus is not represented in the genomes of maize or tobacco since a radiolabelled AtGSHS probe failed to hybridise to genomic DNA of these species (data not shown). High levels of expression of the corresponding mRNA was detected by Northern blot analysis of total RNA isolated from roots and leaves using the *Xho*I fragments of clone pM12B as a radiolabelled probe (Fig. 5). The size of the mRNA detected was 2.5 kb, which is longer than that of the isolated cDNA clone. No further 5′ sequence was obtained following nested pcr amplification of p λYES.

4. Discussion

In this report we describe the isolation of an *Arabidopsis thaliana* cDNA which encodes a protein with a high degree of amino acid sequence identity to the GSHS from *Xenopus laevis* (42%), *Rattus norvegicus* (40%), *Schizosaccharomyces pombe* (40%), and unpublished sequences from *Homo sapiens* (37%) and *Arabidopsis* (100% and 95%). Since suitable DNA probes or antibodies were not available at the outset of this work to screen a plant cDNA library for GSHS sequences we adopted the approach of functional complementation of *E. coli* mutants deficient in this enzyme. We have previously used this approach to isolate and characterise an *Arabidopsis* cDNA encoding γ-glutamylcysteine synthetase [8]. Strong evidence has been obtained to support the hypothesis that the cDNA isolated encodes *Arabidopsis* GSHS. The presence of the plasmid pM12A containing an *Arabidopsis* cDNA in the *E. coli gsh*B mutant 830, which lacks GSHS activity and measurable GSH, leads to restoration of extractable GSH to levels equivalent to those in the wild type, AB1157. A comparison of the derived amino acid sequence of the AtGSHS cDNA clone showed significant homology over the entire sequence to the previously described sequences for GSHS from *X. laevis* [16], *R. norvegicus* [17], *S. pombe* [18] and *H. sapiens*.

Both chloroplastic and cytosolic isoforms of GSHS exist in plants [6]. Whilst the N-terminus of the *Arabidopsis* sequence is longer than that of any other species, we cannot say in which cellular location the encoded peptide functions since there is no

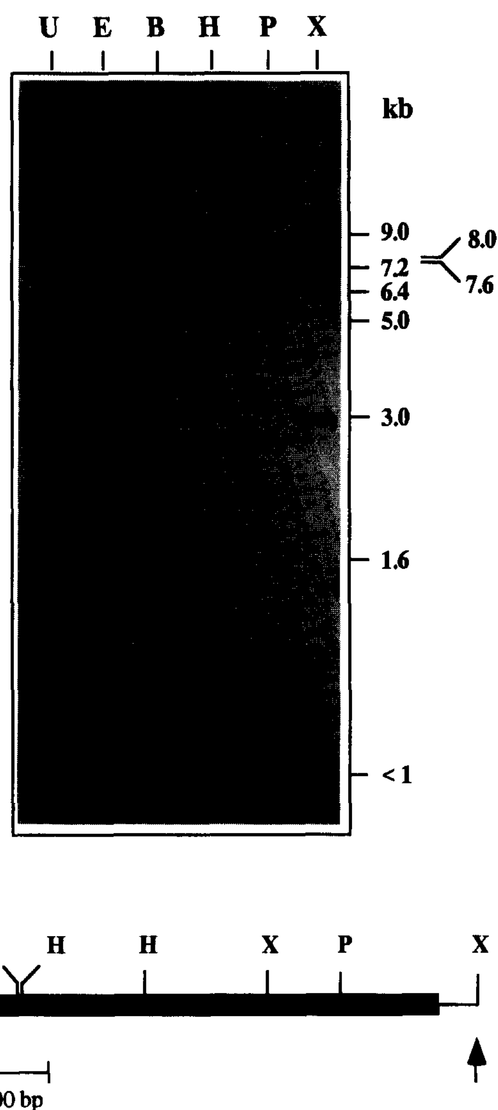

Fig. 4. Genomic Southern analysis of *Arabidopsis GSHS*. *Total Arabidopsis* genomic DNA (20 μg), was digested with the enzymes *Eco*RI (lane E), *BamH*I (lane B), *Hind*III (lane H), PstI (lane P), *Xho*I (lane X) or left uncut (lane U). *Xho*I fragments of pM12B were used together as a probe. Restriction sites within the cDNA are indicated on the above map. Arrows denote the *Xho*I cloning site of pM12A and the open reading frame is represented as a thick black line. The scale is as shown. Fragment sizes are indicated on the right.

Table 2
Homologies of derived amino acid sequences for GSHS

	At1	At2	At3	X1	Rn	Sp	Hs	Ac	Ec
At1	100/100								
At2	100/100	100/100							
At3	97/95	97/95	100/100						
X1	65/42	65/42	64/41	100/100					
Rn	62/40	62/40	62/37	78/64	100/100				
Sp	63/40	63/40	60/37	63/40	62/42	100/100			
Hs	60/37	60/37	60.37	71/57	84/78	60/37	100/100		
Ac	42/16	42/17	42/17	47/18	48/23	43/19	43/15	100/100	
Ec	46/15	44/16	44/15	42/15	48/21	40/18	42/20	59/33	100/100

The values given represent percent similarity followed by percent identity.

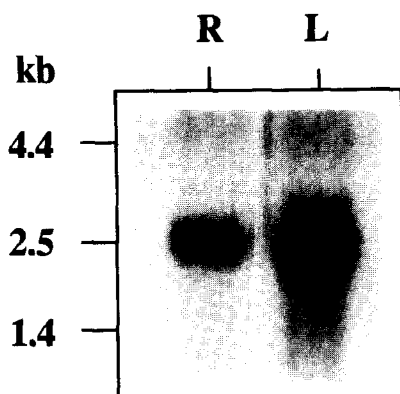

Fig. 5. Expression of GSHS in *Arabidopsis* total RNA (10 μg) from flask grown roots (lane R) and leaves (lane L). *Xho*I fragments of pM12B were used together as a probe. RNA size markers are shown on the left.

stop codon upstream of the first methionine and hence we do not know if the entire N-terminus is represented by the cDNA. It is thus possible that a transit peptide is encoded in the missing 5′ sequence.

Whilst the *Arabidopsis* GSHS sequence described here shares significant homology with the primary structure of comparable eukaryotic sequences, there is very poor homology with published prokaryotic sequences from *E. coli* [19] and *Anaplasma centrale* [20] (16% identity). Clearly two classes of GSHS exist and within each class homology is shared between individual members although clear sub-groups exist; *Arabidopsis* sequences being clearly distinct from the strongly homologous mammalian sequences or yeast sequence (Table 2). Nevertheless, although structural diversity has evolved, GSHSs from all sources share similar functional properties and substrate specificities. A similar situation has been previously described for *Arabidopsis* γ-glutamylcysteine synthetase [8]. Functional properties of the enzymes are therefore likely to be determined by tertiary and quaternary structure. For example it has recently been reported that the tertiary structures of GSHS and D-Ala:D-Ala ligase from *E. coli*, both ATP-hydrolysing proteins, are remarkably similar in spite of insignificant primary sequence homology [21]. Because a single *Arabidopsis* cDNA effectively complemented the *E. coli* mutant we assume that the *Arabidopsis* enzyme functions as a single polypeptide or as a multimer composed of identical subunits, and that the active site is encoded within the cloned sequence. Thus, whilst the *Arabidopsis* enzyme shares strong primary sequence homology with the enzyme from *S. pombe*, this is not reflected at the level of quaternary structure since the *S. pombe* holoenzyme is a heterodimer composed of a 33 and 26 kDa subunit [22]. In contrast, the *R. norvegicus* enzyme is a dimer consisting of two identical subunits of approximately 55 kDa [17], whereas the *A. centrale* enzyme is a single polypeptide of 34 kDa [20] and the *E. coli* enzyme functions as a homotetramer made up of identical 38 kDa subunits [19]. The proposed subunit structure of the *X. laevis* enzyme as a dimer of non-identical subunits [16] must be substantiated since the native enzyme has not yet been isolated. Thus, although there are strong homologies amongst the predicted primary sequence of eukaryotic GSHSs, there is a high degree of diversity in their quaternary structure. Little is known about GSHS from plants and to date there has been only one report of enzyme purification from legumes [23]. Both pea and mung bean enzymes are proteins with an apparent molecular weight of 85 kDa as determined by gel filtration. The cDNA described encodes a protein with a predicted moleclular mass of 58 kDa. However, given the discrepancy between the size of the cDNA and the corresponding mRNA it is likely that the N-terminus extends further and thus the mature protein may be closer in size to that of pea and mung bean. The sequence that is not present in the cDNA may not be essential for enzyme activity since the cDNA clone encodes a protein which fully complemented the GSHS-deficient mutant and restored GSH levels equal to those in the wild type.

The open reading frames contained within the AtGSHS, *X. laevis* and *S. pombe* clones all extend from the beginning of the available nucleic acid sequence and within this region amino acid homologies exist. It is possible that the N terminus of the *X. laevis* and *S. pombe* GSHSs also extend further. It is at present unclear why all of the eukaryotic GSHS clones represented in Fig. 3 have an apparently truncated 5′ end, particularly since all the clones originate from different sources and were obtained by different cloning strategies. Attempts to obtain further *Arabidopsis* 5′ sequence from the original pYES library using pcr based techniques have proven unsuccessful. Resolution of this problem will come through the isolation of genomic GSHS clones. The predicted amino acid sequence of the *X. laevis* GSHS is virtually identical in length to that of the AtGSHS however the cDNA encoding *X. laevis* GSHS contains a 3′ untranslated region of approximately 1000 bp. increasing the total length of the cDNA to 2489 bp [16]. It is likely that there is also further 3′ sequence lacking from the AtGSHS sequence since neither Poly adenosylation signals nor Poly A tail were observed thus explaining in part the discrepency between the size of the cDNA and the size of the mRNA detected by Northern hybridisation.

Genomic Southern blot analysis demonstrated that the gene encoding GSHS, *GSHB*, is represented by a single copy in the *Arabidopsis* genome. This would suggest that little sequence homology may exist between the genes for the cytosolic and chloroplastic isoforms of the enzyme or that they are both encoded by the same gene. Further Southern blot analysis using DNA fragments corresponding to highly conserved regions may reveal sequences corresponding to the second isoform of the enzyme. Lack of homology between isoforms as determined by genomic Southern analysis was also observed for γ-glutamylcysteine synthetase, the first of the two enzymes responsible for the synthesis of glutathione [8]. Similarly, both AtGSHS and At γ-glutamylcysteine synthetase failed to hybridise to maize and tobacco DNA highlighting that within both dicots and monocots, considerable structural diversity of the genes encoding the enzymes in this important biosynthetic pathway has evolved.

The isolation of a cDNA encoding GSHS from *Arabidopsis* is of importance for a number of reasons. Firstly, a plant cDNA encoding GSHS has been isolated by the novel approach of functional complementation, and this coupled with our success in the isolation of a cDNA encoding γ-glutamylcysteine synthetase [8] will allow a rigorous examination of the factors which regulate the glutathione biosynthetic pathway. Secondly, the availability of GSHS cDNA is of practical importance in the construction of transgenic plants, which may have an altered capacity for GSH biosynthesis. Such plants will allow us to

address directly fundamental questions concerning the physiological requirements for GSH in higher plants.

Acknowledgements: We gratefully acknowledge the assistance of John Baker with photography. MJM is supported by a Glasstone research fellowship and an EMBO long-term postdoctoral fellowship. MRR is supported by a grant from The Ministry of Agriculture Fisheries and Food.

References

[1] Halliwell, B. and Gutteridge, J.M.C. (1989) Free radicals in biology and medicine, Oxford, Clarendon Press.
[2] Sies, H. (1993) Eur. J. Biochem. 215, 213–219.
[3] Bowler, C., Van Montagu, M. and Inze, D. (1989) Annu. Rev. Plant Physiol. Plant Mol. Biol. 43, 83–116.
[4] Inze, D. and Van Montagu, M. (1995) Curr. Opin. Biotechnol. 6, 153–158.
[5] Foyer, C.H. and Halliwell, B. (1976) Planta 133, 21–25.
[6] Alscher, R.G. (1989) Physiol. Plant. 77, 457–464.
[7] May, M.J. and Leaver, C.J. (1993) Plant Physiol. 103, 621–627.
[8] May, M.J. and Leaver, C.J. (1994) Proc. Natl. Acad. Sci. USA 91, 10059–10063.
[9] Apontoweil, P. and Berends, W. (1975) Biochim. Biophys. Acta 399, 10–22.
[10] Elledge, S.J., Mulligan, J.T., Ramer, S.W., Spottswood, M. and Davis, R.W. (1991) Proc. Natl. Acad. Sci. USA 88, 1731–1735.
[11] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A Laboratory Manual, Cold Spring Harbor Lab. Press, Plainview, NY.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983) Plant Mol. Biol. Rep. 1, 19–21.
[14] Gurr, S.J. and McPherson, M.J. (1992) in: S.J. Gurr, M.J. McPherson and D.J. Bowles (Eds.), Molecular Plant Pathology: A Practical Approach, IRL, Oxford, Vol. 1, pp. 109–121.
[15] Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (1990) PCR Protocols, London, Academic press.
[16] Habenicht, A., Hille, S. and Knochel, W. (1993) Biochim. Biophys. Acta 174, 295–298.
[17] Huang, C.-S., He, W., Meister, A. and Anderson, M.E. (1995) Proc. Natl. Acad. Sci. USA 92, 1232–1236.
[18] Mutoh, N., Nakagawa, C.W., Ando, S., Tanabe, K. and Hayashi, Y. (1991) Biochem. Biophys. Res. Commun. 181, 430–436.
[19] Gushima, H., Yasuda, S., Soeda, E., Yokota, M., Kondo, M. and Kimura, A. (1984) Nucl. Acids Res. 12, 9299–9307.
[20] Peters, J.M., Dalrymple, B.P. and Jorgensen, W.K. (1992) Biochem. Biophys. Res. Commun. 182, 1040–1046.
[21] Fan, C., Moews, P.C., Shi, Y., Walsh, C.T. and Knox, J.R. (1995) Proc. Natl. Acad. Sci. USA 92, 1172–1176.
[22] Nakagawa, C.W., Mutoh, N. and Hayashi, Y. (1993) Biochem. Cell. Biol. 71, 447–453.
[23] Macnicol, P.K. (1987) Plant Science 53, 229–235.