J Mol Evol (1995) 41:1048–1056

JOURNAL OF MOLECULAR EVOLUTION

Glutathione S-Transferase and S-Crystallins of Cephalopods: Evolution from Active Enzyme to Lens-Refractive Proteins

Stanislav I. Tomarev, Sambath Chung,* Joram Piatigorsky

Laboratory of Molecular and Developmental Biology, National Eye Institute, National Institutes of Health, Bldg. 6, Room 203, 6 Center Dr. MSC, Bethesda, MD 20892-2730, USA

Received: 6 April 1995 / Accepted: 9 June 1995

Abstract. Our previous studies have shown that the S-crystallins of cephalopod (*Ommastrephes sloani pacificus*) eye lenses comprise a family of at least ten members which are evolutionarily related to glutathione S-transferase (GST, EC 2.5.1.18). Here we show by cDNA cloning that there are at least 24 different S-crystallins that are 46–99% identical to each other by amino acid sequence in the squid *Loligo opalescens*. In each species, all but one S-crystallin (SL11 in *O. pacificus* and Lops4 in *L. opalescens*) examined has an inserted central peptide of variable length and sequence. cDNA expression studies conducted in *Escherichia coli* showed that squid GST (which is expressed little in the lens) has very high enzymatic activity using 1-chloro-2, 4-dinitrobenzene (CDNB) as a substrate; by contrast, SL20-1 of *O. pacificus* and Lops12 of *L. opalescens* (which are encoded by abundant lens mRNAs) have no GST activity. Interestingly, SL11 and Lops4 have some enzymatic activity with the CDNB substrate. Site-specific mutations at Y7 or W38, both residues essential for activity of vertebrate GSTs, or insertion of the central peptide present in the inactive SL20-1, reduced the specific activity of squid GST by 30- to 100-fold. These data indicate that the S-crystallins consist of a family of enzymatically inactive proteins (when using CDNB as a substrate) which is considerably larger than previously believed and that GST activity was lost by gradual drift in sequence as well as by insertion of an extra peptide by exon shuffling. The results are also consistent with the idea that SL11 and Lops4 are orthologous crystallins representing the first descendants of the ancestral GST gene in the pathway which gave rise to the extensive S-crystallin family of lens proteins.

Key words: Lens — Squid — Crystallin — Glutathione S-transferase — Parallel evolution

Introduction

Approximately 90% of the water-soluble proteins of the eye lens are crystallins, which are largely responsible for the refractive properties of this transparent tissue (Wistow and Piatigorsky 1988; de Jong et al. 1994). In vertebrates, crystallins are surprisingly diverse and may differ both quantitatively and qualitatively among species. The crystallins that are confined to specific species are called taxon specific and are either related or identical to metabolic enzymes expressed at lower concentrations in many tissues. Consequently they are called enzyme-crystallins (Piatigorsky and Wistow 1989; de Jong et al. 1989). In vertebrates, enzymes have been recruited to become crystallins by being overexpressed in the eye lens, a process which often occurred without gene duplication (i.e., lactate dehydrogenase/ε-crystallin or α-enolase/τ-crystallin) (see Piatigorsky and Wistow 1991; Piatigorsky 1992; Wistow 1993). In some cases a gene duplication has taken place, with one daughter gene becoming specialized for crystallin expression and the

* *Present address:* Department of Biochemistry and Molecular Biology, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854
Correspondence to: S.I. Tomarev

other maintaining its enzymatic function as well as serving as a crystallin (i.e., argininosuccinate lyase/δ-crystallin). One of the most interesting hallmarks of this situation, called gene sharing, consists of the identical protein serving a refractive function in the lens (a crystallin) or an enzymatic function in nonlens tissues depending on the differential expression of its gene.

Although much less is known about crystallins of invertebrates, it is clear that the same strategy of recruiting metabolic enzymes as refractive proteins has been employed in some cases. Cephalopods (squid, octopus, cuttlefish) contain complex eyes with many similarities to the eyes of vertebrates (Packard 1972; Land and Fernald 1992). Such similarities appear to be the product of parallel, rather than convergent, evolution (Zuckerkandl 1994; Zuker 1994). Cephalopods have cellular lenses filled mainly with one major class of soluble protein (S-crystallin; Siezen and Shaw 1982; Chiou 1984) evolutionarily related to glutathione S-transferases (GSTs) (Wistow and Piatigorsky 1987; Tomarev and Zinovieva 1988). Multiple gene duplications have occurred among the S-crystallins of the squid *Ommastrephes pacificus* giving rise to a relatively large family of related proteins (Tomarev et al. 1992). In addition to the GST-related S-crystallins that are expressed highly in the lens and cornea but not in other tissues (Tomarev et al. 1991, 1992; Cuthbertson et al. 1992), cephalopods contain a single GST gene encoding an active enzyme expressed principally in the digestive gland (Harris et al. 1991; Tomarev et al. 1993; Tang et al. 1994). In contrast to the situation among enzyme-crystallins in vertebrates, where the same gene encodes the active enzyme and the crystallin, the GST gene of cephalopods is expressed at very low levels in the lens. Experiments with lens extracts have suggested but not proved that S-crystallins have little if any GST activity (Tomarev et al. 1991, 1992).

The evolutionary process that converts one or more GSTs to S-crystallins is not known. In *O. pacificus* there is only 42–44% identity in amino acid sequence between GST and the S-crystallins, although residues essential for the activity of vertebrate GSTs have been conserved in both the squid GST and S-crystallins (Tomarev and Zinovieva 1988; Tomarev et al. 1992, 1993). Moreover, most of the S-crystallin genes have acquired an extra exon (number 4) which encodes a peptide of variable sequence and length among the different S-crystallins. Interestingly, however, there is at least one S-crystallin (SL11) which lacks exon 4, as does the GST gene.

Here we have characterized by cDNA cloning an extensive family (at least 24 members) of S-crystallins of another species of squid, *Loligo opalescens*. We have directly compared by cDNA expression studies the GST activity of squid GST and S-crystallins, and have tested by mutagenesis experiments the functional importance of certain critical amino acids of GST. The results have led to a model for the evolution of the S-crystallins in cephalopods that involves multiple gene duplications, sequence drift, and exon shuffling.

Materials and Methods

Isolation of Nucleic Acids. RNA was isolated by the acidic guanidinium thiocyanate–phenol–chloroform extraction method (RNazol B, Cinna/Biotecx, Friendswood, TX) from squid *L. opalescens* lenses and stored at −70°C. Poly(A)$^+$RNA was prepared using the Dynabeads mRNA purification kit (Dynal Inc., Oslo, Norway). Plasmid DNAs were isolated using the Wizard DNA purification kit (Promega, Madison, WI)

Construction and Screening of a Lens cDNA Library. A lens cDNA library was constructed from *L. opalescens* as described previously (Tomarev et al. 1992) using the ZAP II vector (Stratagene, La Jolla, CA). The library contained about 10^6 independent recombinants. cDNA clones encoding different S-crystallins (Lops clones) were isolated by screening the library with labeled probes prepared by polymerase chain reaction (PCR) of cDNA synthesized using *L. opalescens* lens poly(A)$^+$RNA as template and oligonucleotides (T/C)TGTA(T/C)TACTTCAA(T/C)(A/G)G(C/T)CG(T/C)GG(A/G)CG(T/G)GC and $(T)_{30}$ as primers. The first (5′) primer corresponds to a conserved amino acid sequence (positions 5–14) among cephalopod S-crystallins (Tomarev and Zinovieva 1988; Tomarev et al. 1991, 1992). cDNA inserts were sequenced by the dideoxynucleotide termination method using Sequenase Version 2.0 (Amersham). Sequences were determined for both strands of DNA and analyzed using the GCG program package (Devereux et al. 1984).

Expression of Squid GST and S-crystallins in E. coli. The cDNA inserts of pGST5 (Tomarev et al. 1993) encoding squid GST, of pSL11 and pSL20 (Tomarev and Zinovieva 1988) encoding *O. pacificus* S-crystallins, and of Lops4 and Lops12 (see Fig. 1) encoding *L. opalescens* S-crystallins were amplified by PCR using Vent polymerase (New England Biolabs, Beverly, MA). The primers used created an *Nde*I restriction site immediately 5′ to the translation initiation codon and a *Not*I site in the 3′ untranslated region of the amplified cDNAs. After digestion with *Nde*I and *Not*I the PCR products were fractionated by agarose electrophoresis and those with the expected lengths were cut out of the gel, purified using Geneclean (BIO 101, La Jolla, CA), ligated in the expression vector pET-17b (Novagen, Madison, WI), and used to transform *E. coli* strains DH5α or BL21(DE3). The plasmids were isolated and their cDNA inserts were sequenced. Constructs which were grown initially in DH5α were regrown in BL21(DE3). Point mutations were introduced by the method of Jones and Howard (Jones and Howard 1990) and insertions and deletions were performed as described by Ho et al. (1989). Vent polymerase was used for PCR in both cases. Bacteria containing recombinant constructions were grown in LB medium supplemented with ampicillin to an optical density 0.6–0.8, induced by isopropylthio-β-galactoside (IPTG), and incubated for another 3 h at 30°C (induction at 37°C resulted in the accumulation of the expressed proteins within inclusion bodies). Bacterial cells were collected by centrifugation, washed once in 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, and resuspended in the same buffer in one-tenth the original volume. Cells were disrupted by sonication and the insoluble fraction was removed by centrifugation at 12,000*g* for 15 min at 4°C. The unfractionated cell lysate, cytosol, and pellet were analyzed by SDS-PAGE and the cytosolic fraction was used for GST assays. The relative amount of recombinant proteins was calculated after scanning the gels with SciScan 5000 (U. S. Biochemicals).

GST Assay. E. coli extracts and purified GST were assayed for GST activity with CDNB as substrate as described (Habig et al. 1974).

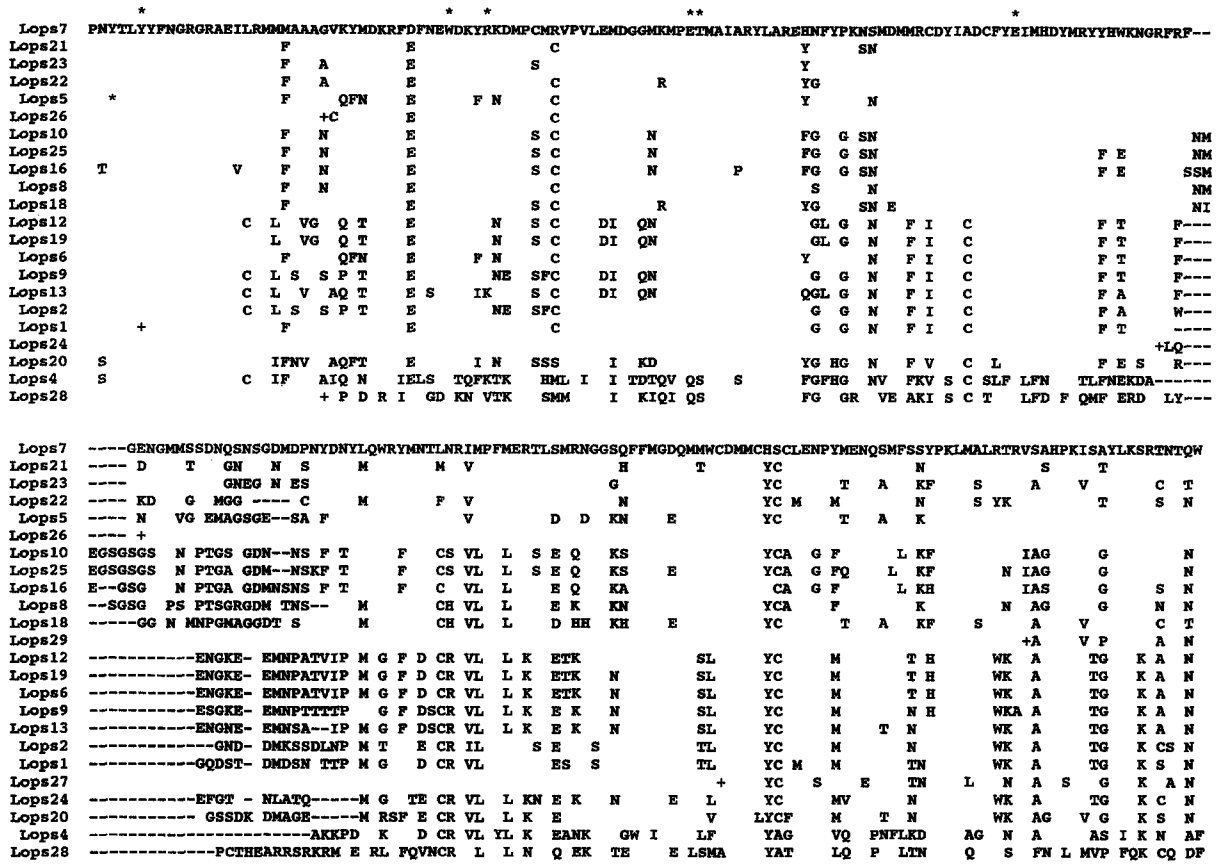

Fig. 1. Comparison of amino acid sequences of squid (*L. opalescens*) S-crystallins encoded by Lops1–Lops29. Sequences were aligned using the PileUp program (Devereux et al. 1984). Sequence of S-crystallin encoded by Lops7 is shown in full; for other sequences only differing amino acids are shown. *Dashes* show the gaps which were introduced to maximize similarity. + marks the beginning and the end of incomplete sequences. * marks the residues whose counterparts in the squid GST are essential for GSH binding. The nucleotide sequences reported in this paper has been submitted to the GenBank/EMBL Data Bank with accession numbers U19255–U19300, U19387–U19390.

Specific activities of GST, different S-crystallins, and their mutants were measured with 1 mM glutathione (GSH) and 1 mM CDNB. Kinetic constants for CDNB and GSH were measured with 1 mM GSH or 1 mM CDNB and varying concentrations of the other substrate, ranging usually from 0.1 to 2 mM.

Phylogenetic Analysis. Aligned amino acid sequences of S-crystallins with deleted central inserts were analyzed by the PHYLIP package (version 3.5 by Felsenstein 1993). The SEQBOOT program was used to create multiple (n = 100) randomly sampled data sets from the original alignment. The maximum parsimony program PROTPARS was used on multiple data sets. The program CONSENSE was used to obtain a consensus tree as well as confidence levels. Aligned nucleotide sequences of S-crystallin cDNAs after deletion of the regions encoding the central inserts and untranslated sequences were analyzed similarly using the SEQBOOT, DNAPARS, and CONSENSE programs.

Results

Characterization of the S-Crystallin Family of L. opalescens

Nineteen novel S-crystallin cDNAs were isolated from an *L. opalescens* library after screening with a PCR probe derived from total RNA using a degenerate 5′ primer from a conserved region of *O. pacificus* S-crystallins and a poly(dT) 3′ primer (see Materials and Methods). The deduced amino acid sequences of these cDNAs are given in Fig. 1. Four other partial cDNAs encoding different S-crystallins (Lops25–Lops29) were also identified during our attempts to clone transcription factors from squid lens by PCR. The cDNAs encoding S-crystallins are 58–98% identical to each other in nucleotide sequence and their deduced proteins are 46–99% identical in amino acid sequence. *L. opalescens* S-crystallins show 36–44% identity to the GST from *O. pacificus*. The most variable regions of these S-crystallins are situated centrally, as is the case with the S-crystallins of *O. pacificus* (Tomarev and Zinovieva 1988; Tomarev et al. 1992). In *O. pacificus* this variable sequence is encoded by a separate exon (number 4) in each gene that has been examined (Tomarev et al. 1992). Only one *O. pacificus* S-crystallin gene (SL11) lacking this variable exon 4 has been identified. Similarly, only one S-crystallin (Lops4) without a central insert was identified in *L. opalescens*.

Phylogenetic trees were constructed using the maximum parsimony method for both DNA and deduced pro-

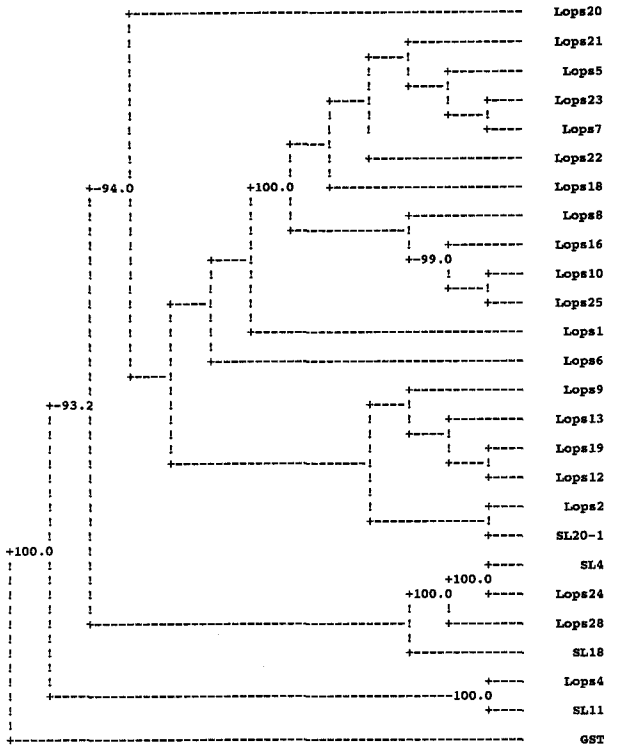

Fig. 2. Maximum parsimony phylogenetic tree of the squid GST and S-crystallins of two species of squid. *L. opalescens* S-crystallins (*Lops*) are from this work. *O. pacificus* S-crystallins and GST are from our previous work (Tomarev and Zinovieva 1988; Tomarev et al. 1992, 1993). This is an unrooted tree; GST was arbitrarily chosen as outgroop root. Bootstrap values of 90% and higher (from 100 replicates) are indicated. Branch lengths are arbitrary.

tein sequences. Sequences corresponding to the central inserts were not included, since the acquisition of these sequences within the family could take place independently several times in the course of evolution (see below). The general topology of the trees was similar for the protein and nucleotide sequences and thus only the parsimony tree for protein sequences is shown (Fig. 2). Although the bootstrap numbers were not high for some S-crystallin nodes, the two S-crystallins which naturally lack the central insert (SL11 and Lops4) always formed a separate group and clustered together with GST. SL11 and Lops 4 are 83% identical in amino acid sequence, and both have relatively low methionine contents (5%); by contrast, the other S-crystallins are methionine-rich (10–14%). The relationships among the other S-crystallins in *O. pacificus* and *L. opalescens* are less clear.

Two crystallins, SL20-1 in *O. pacificus* and Lops12 in *L. opalescens,* were encoded by mRNAs whose sequences were especially abundant in the lens cDNA libraries. SL20-1 and Lops12 are 80% identical in their overall amino acid sequences; this identity reaches 86% when the variable central peptide sequence is not included. Thus, the S-crystallin family in cephalopods is considerably larger than believed previously (Tomarev et al. 1992) and consists of at least 24 members in *L. opalescens.*

```
                *                                   *   *
GST      PKYTLHYFPLMGRAELCRFVLAAHGEEFTDRVVEMADWPNLKATMYSNAM   50
GST MUT        F                                     F     HML
SL20-1    N   Y  NGR    I  MLM  A VQY  KRF FNE DKYRND P MCV   50
Lops12    N   Y  NGR    I  MLM VG VQY  KRF FNE DKYRND P MCV   50
SL11      S   Y  NGR    I  MLF VASVQYQ KRI L E TQF TK PCHML   50
Lops4     S   Y  NGR    I  MIF  AAIQYN K   LSE TQF TK PCHML   50

                  **                                    *
GST      PVLDIDG-TKMSQSMCIARHLAREFGLDGKTSLEKYRVDEITETLQDIFN   99
SL20-1         QN  PET A   Y    N YY  NNMDMF I Y CDCFYE LH   100
Lops12         QN  PET A   Y    H  Y  NNMDMF I Y CDCFYE MH   100
SL11      I E  TE QVP   A S Y      FY  NNMDMFK  CLCDS FEL    100
Lops4     I E  TD QVP   A S Y      FH  NNVDMFK  S CDS FEL    100

GST      DVVKIKFAPEAAK-----------------EAVQQNYEKSCKRLAPFLE   131
GST MUT   YMA          MQGSGTDMSPDMDPTQMT
SL20-1    YMRYFETKNGRFMQGSGTDMSPDMDPTQMTSYI NR LDT R IL       150
Lops12    YMRYFETKNGRFFENGKESEMNPATVI---PYM GRFMDT R VL       147
SL11      YMAVYNEKD   K-----------------TEL KRFQNT L VL YM    133
Lops4     YMTLFNEKD   K-----------------PDL KR MDT R VL Y     133

GST      GLLVSNGGGDGFFVGNSMTLADLHCYVALEVPLKHTPELLKDCPKIVALR   181
SL20-1    RT EMRN  KE  M DQ M C MM  CC  N MLEDQTTFNNF  LMS W   200
Lops12    KT ETKN  SQ  M DQ S C MM  CC  N MMENQSMFTSH  LM  W   197
SL11      KT EA K  A W I DQIL C MMTHA   N IQENAN   EY  LA      183
Lops4     KT EA K  S W I DQ LFC MM  AG  N VQEN NF   Y  LAG     183

GST      KRVAECPKIAAYLKKRPVRDF                               202
SL20-1       SH   TP      NNTNW                              221
Lops12       AH   TG      ANTNW                              218
SL11      T  AH       I   NNTA                               204
Lops4     N  AH    S I    NNTA                               204
```

Fig. 3. Amino acid alignment of GST (Tomarev et al. 1993), SL20-1, and SL11 (Tomarev and Zinovieva 1988) of *O. pacificus* and Lops4 and Lops12 of *L. opalescens.* The GST sequence is shown in full; only the differing amino acids are shown for the S-crystallins. *Dashes* show the gaps which were introduced to maximize similarity. * marks the residues essential for GSH binding. *GST MUT* shows individual mutations which were introduced into the squid GST.

GST activity of expressed proteins in E. coli

Previous experiments have shown that purified GST from the digestive gland of *O. pacificus* is extremely active enzymatically while S-crystallins in lens extracts have very little enzymatic activity using CDNB as a substrate (Tomarev et al. 1993). Those experiments, however, did not distinguish among the multiple S-crystallins in the lens extracts. Moreover, they did not distinguish between low or negligible GST activity in the primary S-crystallins and inactivation of GST activity by posttranslational modifications of the primary gene products or the presence of an inhibitor in the extract in which the tests were conducted. We have thus expressed the cDNA encoding the active GST from the digestive gland of *O. pacificus* (Tomarev et al. 1993) and four cDNAs encoding S-crystallins from the lens in *E. coli* and assayed their GST activity using CDNB as a substrate. SL20-1 containing a central insert and SL11 lacking a central insert from *O. pacificus* (Tomarev et al. 1992), and cDNAs encoding proteins with similar properties, Lops12 and Lops4, respectively, from *L. opalescens* were chosen for these experiments (see Fig. 3). The principal reason we chose SL20-1 and Lops12 among the multiple S-crystallins containing central inserts was their abundant representation in the cDNA libraries, indicating that they are major crystallins in the lens.

The recombinant proteins comprised 10–20% of the total water-soluble protein after expression in *E. coli,* as judged by SDS-PAGE (not shown). The recombinant

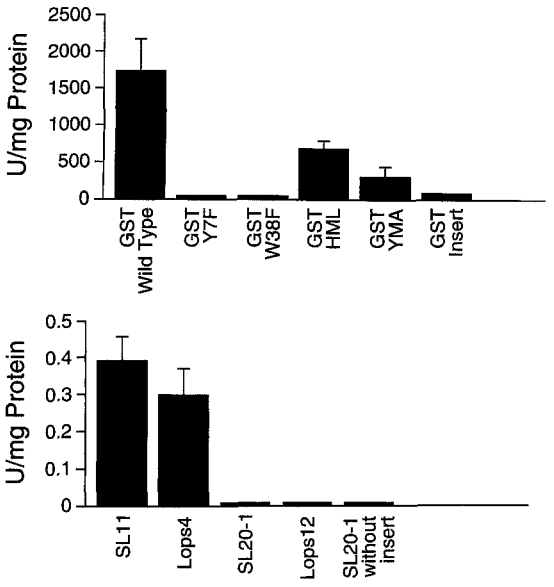

Fig. 4. Specific activities of recombinant wild-type and mutant squid GST and several S-crystallins.

squid GST possessed twice the specific activity using CDNB as substrate as that observed previously (Tomarev et al. 1993) with GST purified from the digestive gland (Fig. 4). The K_m values were similar for the recombinant GST assayed in the *E. coli* extract (Table 1) and that purified from the extract by S-hexylglutathione chromatography (data not shown, see also Ji et al. 1995). Thus, all subsequent experiments on measurement of GST activity were conducted in the *E. coli* extracts. In contrast to the high GST activity of the protein expressed from the digestive gland cDNA, there was essentially no GST activity by the proteins expressed from the SL20-1 and Lops12 S-crystallin cDNAs using the CDNB substrate (Fig. 4). However, the SL11 and Lops4 cDNAs generated proteins with low but clearly detectable GST activity (Fig. 4).

Mutational Analysis of GST

We next tested the possibility that the central insert inactivates SL20-1, since this peptide insert is one of the main structural differences between SL20-1 and SL11 (Tomarev et al. 1992), which has some GST activity. Two constructs were made. First, the nucleotide sequence encoding the insert peptide (which is equivalent to exon 4) was deleted from the SL20-1 cDNA, and second, the insert nucleotide sequence was placed in its homologous position within the squid GST cDNA; both constructs were expressed in *E. coli* and the resulting mutated proteins were assayed for enzymatic activity. As expected, the inserted peptide decreased the activity of GST about 30-fold (Fig. 4). Removal of the insert from SL20-1, however, did not generate enzymatic activity in the shortened S-crystallin (Fig. 4).

We made several site-specific mutations in squid GST (see Fig. 3, GST MUT) in order to compare the structural and functional similarities between the cephalopod and vertebrate enzymes. Studies of others have implicated several residues for enzymatic activity of mammalian GSTs (see Armstrong 1991, 1994; Dirr et al. 1994a; Wilce and Parker 1994, for review). One of these residues is tyrosine 7 (Y7), which is essential for catalytic activity but not for GSH binding (Sternberg et al. 1991; Wang et al. 1992; Kolm et al. 1992; Kong et al. 1992; Liu et al. 1992). Another critical residue is tryptophan 38 (W38), which is necessary for both catalytic activity and GSH binding (Reinemer et al. 1991; Nishihira et al. 1992; Baker et al. 1994). These are conserved in the squid GST (Fig. 3, asterisks). Changing Y7 to phenylalanine (Y7F) or W38 to phenylalanine (W38F) decreased the enzymatic activity of squid GST 50- to 100-fold (Fig. 4), while the K_m for GSH was higher only for the W38F mutation (Table 1). In another mutation, VVK at positions 101–103 of GST was changed to YMA, as in SL11, since these amino acids are believed to lie in the substrate binding site of mammalian GST (Johnson et al. 1993; Ji et al. 1994; Dirr et al. 1994a,b). This mutation resulted in an approximate sevenfold loss of GST activity (Fig. 4). Finally, NAM was changed to HML, also as in SL11, at positions 48–50 in squid GST, since computer modeling based on the X-ray structure of mammalian class mu GST (Ji et al. 1992) suggested that these three amino acids may lead to different conformations between GST and SL11.[1] The NAM to HML change led to an approximate threefold loss of GST activity (Fig. 4). Taken together these results are consistent with significant structural and functional similarities between squid and vertebrate GSTs.

Table 1. K_m values (m*M*) of the squid wild-type and mutant GST and S-crystallins

Protein	GSH K_m	CDNB K_m
GST	1.4 ± 0.3	1.2 ± 0.4
GST Y7F	0.66 ± 0.12	0.7 ± 0.1
GST W38F	3.7 ± 0.2	2.8 ± 0.2
GST HML	2.3 ± 0.5	1.5 ± 0.3
GST YMA	1.1 ± 0.2	1.8 ± 0.4
SL11	0.36 ± 0.12	1.7 ± 0.4
Lops4	0.33 ± 0.08	6.3 ± 0.8

Discussion

Previous studies have shown that the GST-related S-crystallins of *O. pacificus* comprise a relatively large family of at least ten different proteins (Tomarev et al.

[1] Wistow, G.—personal communication.

1992). In the present investigation we show that the family of S-crystallins may be considerably larger: at least 24 members are present in the squid *L. opalescens*. One S-crystallin pseudogene has also been isolated from this species (data not shown). In view of the appreciable differences in sequence among the related S-crystallins (46–99% identical), it is likely that the S-crystallins have been progressively expanding during evolution, some members having appeared much more recently than others by gene duplication. It is possible that different S-crystallins have different structural and/or functional roles in the lens, accounting for their differences in sequence. Possibly the multiplicity and variability of S-crystallins prevent their crystallization due to their elevated concentration in the lens (which exceeds that of crystallins in vertebrates) and high hydrostatic pressures generated under water (Gross and Jaenicke, 1994).

The present cDNA expression data establish that S-crystallins have no (SL20-1/Lops12) or little (SL11/Lops4) GST activity. The three-dimensional structure of squid GST was resolved recently (Ji et al. 1995). Sequence alignment with the squid GST suggests that SL11 and Lops4 crystallins retain most of the residues that participate in the binding of GSH. Of the six residues (Y7, W38, K42, Q62, S63, and D96, see Fig. 3, asterisks) of GST that have side chains directly involved in GSH binding, five have identical counterparts in SL11 and Lops4. The only exception is E97 of SL11/Lops4, which corresponds to D96 of GST. There are three additional mutations in the residues essential for GSH binding in SL20-1/Lops12 (K42, Q62, and S63 of GST are replaced by R42, E63, and T64, respectively); indeed, these are observed in all S-crystallins which we analyzed (see Fig. 1, asterisks). The only exceptions to these amino acid replacements among the S-crystallins are Lops28 where K42 has been changed to V42 and Lops13 where K42 has not been changed. These amino acid differences between GST and S-crystallins are consistent with the latter failing to bind to a GSH or a S-hexylglutathione affinity column. GSH binding to S-crystallins would probably be deleterious to the lens, since it would decrease the availability of GSH for protection against oxidative damage which could lead to cataract formation (Spector 1991). While it remains possible that the S-crystallins would be enzymatically active if tested on a different substrate, we believe that the extremely low GST enzymatic activities in lens extracts determined earlier (Tomarev et al. 1993) cannot be attributed to the presence of an inhibitor or to posttranslational modifications of the primary gene products.

The fact that the enzymatically active S-crystallin lacks the peptide insert present in the enzymatically inactive S-crystallins indicates that introduction of the peptide by exon shuffling (Tomarev et al. 1992) was one of the mechanisms by which S-crystallins lost GST activity during evolution. This idea is supported by our demonstration that insertion of the peptide into active GST abolishes GST activity. However, removal of the peptide from SL20-1 did not activate GST activity, confirming (see above) that additional sequence changes, presumably refining its structural role as a crystallin, led to further modifications incompatible with enzymatic activity. Alignment of the squid S-crystallin and GST sequences suggests that the variable insert peptides, encoded by exon 4 in SL20-1, form a loop between α-helical regions 4 and 5 on the surface of the proteins. Insertion of surface loops is one of the most efficient ways to obtain new properties for proteins with a relatively low danger of inducing instability in internal structural framework (El Hawrani et al. 1994). The appreciable variabilities in sequence and length of the central peptides among the different S-crystallins (Fig. 1; Tomarev et al. 1991; 1992) make it likely that their insertions occurred independently by exon shuffling at different times as this gene family multiplied during evolution. It is possible of course that insertion occurred only once, and the resulting peptide was not very highly constrained by selective pressures. We speculate that multiple insertion events took place by unequal crossing over driven by tandem repetitive sequences (unpublished) which are present in the third intron of the SL20-1 gene (which precedes exon 4 encoding the insert peptide) and of the SL11 gene (which lacks an insert peptide). Single copies of this repetitive sequence are also present in intron 4 of the squid GST gene, intron 1 of the SL20-1 gene, and the 5′ flanking region of the SL11 gene. Tandem repetitive sequences in introns are believed to facilitate insertions or deletions by unequal crossing over (Stoppa-Lyonnet et al. 1990; Olds et al. 1993).

A model proposing the process by which GST was recruited to become a family of S-crystallins in cephalopods is diagrammed in Fig. 5. We presume that SL11 and Lops4, which lack a central insert, retain some GST activity and form a separate group on the phylogenetic tree, are orthologous proteins. The model proposes the duplication of an ancestral gene encoding an active GST enzyme, with one of the daughter genes (SL11/Lops4) becoming highly expressed in the lens. Further duplications of the SL11/Lops4 gene gave rise to an extensive family of S-crystallins whose expression became lens-specific. Both sequence changes (slash marks in Fig. 5) and insertions of a central peptide by exon shuffling (*Ins box* in Fig. 5) led to further specializations for crystallin function and loss of enzymatic activity. This differs from the gene-sharing mechanism used for recruiting enzyme-crystallins in vertebrates where the gene encoding the active enzyme is highly expressed in the lens without losing its expression or original function in other tissues (Piatigorsky and Wistow 1989; de Jong et al. 1989; Piatigorsky and Wistow 1991; Piatigorsky 1992; Wistow 1993). In our squid model, the ancestral GST gene never became highly expressed in the lens; it is possible, of course, that it did so and subsequently lost its high lens

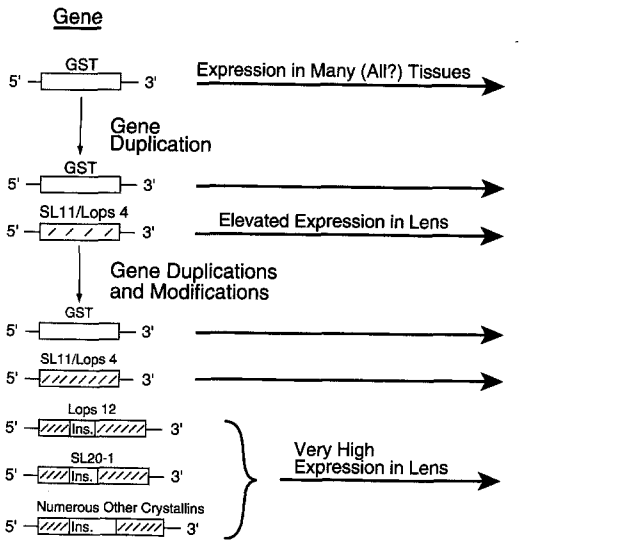

Fig. 5. Recruitment of GST as lens S-crystallins in cephalopods. Ancestral GST gene was expressed in many (all?) tissues. After duplication of this ancestor gene, regulatory elements of one of the duplicated genes acquired the ability for high expression in the lens, and the corresponding protein became an enzyme-crystallin with GST activity. Subsequent duplications of GST/S-crystallin gene accompanied by insertion of the extra exon created the modern family of S-crystallins. *Slash marks,* sequence changes; *Ins box,* insertion of central peptide by exon shuffling.

expression. Thus, separation of function appears to have become complete among most of the S-crystallins of cephalopods, possibly representing a further stage of the evolutionary process in progress among many of the vertebrate enzyme-crystallins.

One of the major challenges for future research is to understand the molecular basis for the high lens-specific expression of the S-crystallins. Sequence and functional comparisons of the promoter regions of five S-crystallin genes, one S-crystallin pseudogene, and the GST gene have provided evidence that the acquisition of an AP-1 binding site in the vicinity of the TATA box has contributed to the high expression of S-crystallins in the lens (Tomarev et al. 1992, 1994). AP-1 sites are prevalent among crystallin genes of vertebrates and appear to have played a significant role in their recruitment (see Piatigorsky and Zelenka 1992 for references). Although GST has not been used as an enzyme-crystallin in vertebrates, the process of recruiting preexisting proteins (generally enzymes) as crystallins is clearly common to vertebrates and cephalopods. This is consistent with the possibility that early eye morphogenesis is under similar genetic control in vertebrates and invertebrates (see Zuckerkandl 1994), with Pax-6 serving as a master control gene (Hill et al. 1991; Ton et al. 1991; Quiring et al. 1994; Zuker 1994; Halder et al. 1995). With respect to crystallins, Pax-6 has been implicated in lens expression of the chicken (Cvekl et al. 1994) and mouse αA (Cvekl et al. 1995a)-, chicken δ1 (Cvekl et al. 1995b)-, and guinea pig ζ (Richardson et al. 1995)-crystallin genes. A potential Pax-6 binding site has been identified in the SL20-1 (positions −44/−23) and SL11 (positions −78/−57 and −72/−93) crystallin genes.[2] These observations support the ideas that candidate crystallin genes were selected not only by the properties of their encoded proteins but also by the nature of their transcription factors (Piatigorsky 1992) and that homology in morphological features between vertebrate and cephalopod lenses reflects a similarity in the networks of regulatory genes (Zuckerkandl 1994). We anticipate that our identification of SL20-1/Lops12 as a nonenzymatically active end-product in S-crystallin recruitment, of SL11/Lops4 as a partially active putative intermediate in this recruitment process, and of squid GST as a very active noncrystallin related to the S-crystallin family will be important for eventually unravelling the molecular mechanisms that led to the conversion of GST to S-crystallins in cephalopods.

Acknowledgments. We thank Drs. W. Gilly for providing laboratory space and help with the collection of the squid tissues, R. Armstrong for providing three-dimensional coordinates of GST 3-3 isozyme, G. Wistow for preliminary three-dimensional modeling of SL11 crystallin, and A. Cvekl for pointing out the potential PAX-6 binding sites in the squid S-crystallin promoters. We thank Frederick Biomedical Supercomputer Center, NCI-FCRDC for allocation of computer time and for technical support.

References

Armstrong RN (1991) Glutathione S-transferases: reaction mechanism, structure, and function. Chem Res Toxicol 4:131–140

Armstrong RN (1994) Glutathione S-transferases: structure and mechanism of an archetypical detoxification enzyme. Adv Enzymol Relat Areas Mol Biol 69:1–44

Baker RT, Smith SA, Marano R, McKee J, Board PG (1994) Protein expression using cotranslational fusion and cleavage of ubiquitin. Mutagenesis of the glutathione-binding site of human Pi class glutathione S-transferase. J Biol Chem 269:25381–25386

Chiou S-H (1984) Physicochemical characterization of a crystallin from the squid lens and its comparison with vertebrate lens crystallins. J Biochem (Tokyo) 95:75–82

Cuthbertson RA, Tomarev SI, Piatigorsky J (1992) Taxon-specific recruitment of enzymes as major soluble proteins in the corneal epithelium of three mammals, chicken and squid. Proc Natl Acad Sci USA 89:4004–4008

[2] Cvekl, A.—personal communication.

Cvekl A, Sax CM, Bresnick EH, Piatigorsky J (1994) A complex array of positive and negative elements regulates the chicken αA-crystallin gene: involvement of Pax-6, USF, CREB and/or CREM, and AP-1 proteins. Mol Cell Biol 14:7363–7376

Cvekl A, Kashanchi F, Sax CM, Brady JN, Piatigorsky J (1995a) Transcriptional regulation of the mouse αA-crystallin gene: activation dependent on cyclic AMP-responsive element (DE1/CRE) and a Pax-6-binding site. Mol Cell Biol 15:653–660

Cvekl A, Sax CM, Li X, McDermott JB, Piatigorsky J (1995b) Pax-6 and lens-specific transcription of the chicken δ1-crystallin gene. Proc Natl Acad Sci USA 92:4681–4685

de Jong WW, Hendriks W, Mulders JWM, Bloemendal H (1989) Evolution of eye lens crystallins: the stress connection. Trends Biochem Sci 14:365–368

de Jong WW, Lubsen NH, Kraft HJ (1994) Molecular evolution of the eye lens. Progr Retinal Eye Res 13:391–442

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Dirr H, Reinemer P, Huber R (1994a) X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. Eur J Biochem 220:645–661

Dirr H, Reinemer P, Huber R (1994b) Refined crystal structure of porcine class pi glutathione S-transferase (pGST P1-1) at 2.1 A resolution. J Mol Biol 243:72–92

El Hawrani AS, Moreton KM, Sessions RB, Clarke AR, Holbrook JJ (1994) Engineering surface loops of proteins—a preferred strategy for obtaining new enzyme function. Trends Biotech 12:207–211

Felsenstein J (1993) PHYLIP (phylogeny inference package) version 3.5c. Department of Genetics, University of Washington, Seattle

Gross M, Jaenicke R (1994) Proteins under pressure. The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. Eur J Biochem 221:617–630

Habig WB, Pabst MJ, Jakoby WB (1974) Glutathione S-transferases. The first step in mercapturic acid formation. J Biol Chem 249: 7130–7139

Halder G, Callaerts P, Gehring WJ (1995) Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila.* Science 267: 1788–1792

Harris J, Coles B, Meyer DJ, Ketterer B (1991) The isolation and characterization of the major glutathione S-transferase from the squid *Loligo vulgaris.* Comp Biochem Physiol 98B:511–515

Hill RE, Favor J, Hogan BLM, Ton CCT, Saunders GF, Hanson IM, Prosser J, Hastie ND, van Heynigen VV (1991) Mouse *small eye* results from mutations in paired-like homeobox-containing gene. Nature 354:522–525

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59

Ji X, Zhang P, Armstrong RN, Gilliland GL (1992) The three-dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-A resolution. Biochemistry 31:10169–10184

Ji X, Johnson WW, Sesay MA, Dickert L, Prasad SM, Ammon HL, Armstrong RN, Gilliland GL (1994) Structure and function of the xenobiotic substrate binding site of a glutathione S-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(S-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene. Biochemistry 33:1043–1052

Ji X, von Rosenvinge EC, Johnson WW, Tomarev SI, Piatigorsky J, Armstrong RN, Gilliland GL (1995) Three-dimensional structure, catalytic properties and evolution of a sigma class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods. Biochemistry 34:5317–5328

Jones DH, Howard BH (1990) A rapid method for site-specific mutagenesis and directional subcloning by using the polymerase chain reaction to generate recombinant circles. Biotechnique 8:178–183

Johnson WW, Liu S, Ji X, Gilliland GL, Armstrong RN (1993) Tyrosine 115 participates both in chemical and physical steps of the catalytic mechanism of a glutathione S-transferase. J Biol Chem 268:11508–11511

Kolm RH, Sroga GE, Mannervik B (1992) Participation of the phenolic hydroxyl group of Tyr-8 in the catalytic mechanism of human glutathione transferase P1-1. Biochem J 285:537–540

Kong K-H, Nishida M, Inoue H, Takahashi K (1992) Tyrosine-7 is an essential residue for the catalytic activity of human class Pi glutathione S-transferase: chemical modification and site-directed mutagenesis studies. Biochem Biophys Res Commun 182:1122–1129

Land MF, Fernald RD (1992) The evolution of eyes. Annu Rev Neurosci 15:1–29

Liu S, Zhang P, Ji X, Johnson WW, Gilliland GL, Armstrong RN (1992) Contribution of tyrosine 6 to the catalytic mechanism of isoenzyme 3-3 of glutathione S-transferase. J Biol Chem 267:4296–4299

Nishihira J, Ishibashi T, Sakai M, Nishi S, Kumazaki T (1992) Evidence for the involvement of tryptophan 38 in the active site of glutathione S-transferase P. Biochem Biophys Res Commun 185: 1069–1077

Olds RJ, Lane DA, Chowdhury V, De Stefano V, Leone G, Thein SL (1993) Complete nucleotide sequence of the antithrombin gene: evidence for homologous recombination causing thrombophilia. Biochemistry 32:4216–4224

Packard A (1972) Cephalopods and fish: the limits of convergence. Biol Rev 47:241–307

Piatigorsky J (1992) Innovation associated with changes in gene regulation. J Biol Chem 267:4277–4280

Piatigorsky J, Wistow G (1989) Enzyme/crystallins: gene sharing as an evolutionary strategy. Cell 57:197–199

Piatigorsky J, Wistow G (1991) The recruitment of crystallins: new functions precede gene duplication. Science 252:1078–1079

Piatigorsky J, Zelenka PS (1992) Transcriptional regulation of crystallin genes: cis elements, trans-factors, and signal transduction systems in the lens. Adv Dev Biochem 1:211–256

Quiring R, Walldorf U, Kloter U, Gehring WJ (1994) Homology of the *eyeless* gene of *Drosophila* to the *Small eye* gene in mice and *Aniridia* in humans. Science 265:785–789

Reinemer P, Dirr HW, Ladenstein R, Schaffer J, Gallay O, Huber R (1991) The three-dimensional structure of class π glutathione S-transferase in complex with glutathione sulfonate at 2.3 A resolution. EMBO J 10:1997–2005

Richardson J, Cvekl A, Wistow G (1995) Pax-6 is essential for lens-specific expression of ζ-crystallin. Proc Natl Acad Sci USA 92:4676–4680

Siezen RJ, Shaw DC (1982) Physicochemical characterization of lens proteins of the squid *Nototodarus gouldi* and comparison with vertebrate crystallins. Biochim Biophys Acta 704:304–320

Spector A (1991) The lens and oxidative stress. In: Sies H (ed) Oxidative stress: oxidants and antioxidants. Academic Press, New York, pp 529–558

Stenberg G, Board PG, Mannervik B (1991) Mutation of an evolutionary conserved tyrosine residue in the active site of human class alpha glutathione transferase. FEBS Lett 293:153–155

Stoppa-Lyonnet D, Carter PE, Meo T, Tosi M (1990) Cluster of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangement. Proc Natl Acad Sci USA 87:1551–1555

Tang S-S, Lin C-C, Chang, G-G (1994) Isolation and characterization of octopus hepatopancreatic glutathione S-transferase. Comparison of digestive gland enzyme with lens S-crystallin. J Protein Chem 13:609–618

Tomarev SI, Zinovieva RD (1988) Squid major lens polypeptides are homologous to glutathione S-transferase subunits. Nature 336:86–88

Tomarev SI, Zinovieva RD, Piatigorsky J (1991) Crystallins of the

octopus lens. Recruitment from detoxification enzymes. J Biol Chem 266:24226–24231

Tomarev SI, Zinovieva RD, Piatigorsky J (1992) Characterization of squid crystallin genes. Comparison with mammalian glutathione S-transferase genes. J Biol Chem 267:8604–8612

Tomarev SI, Zinovieva RD, Guo K, Piatigorsky J (1993) Squid glutathione S-transferase. Relationships with other glutathione S-transferases and S-crystallins of cephalopods. J Biol Chem 268:4534–4542

Tomarev SI, Duncan MK, Roth HJ, Cvekl A, Piatigorsky J (1994) Convergent evolution of crystallin gene regulation in squid and chicken: the AP-1/ARE connection. J Mol Evol 39:134–143

Ton CCT, Hirvonen H, Miwa H, Weil MM, Monaghan P, Jordan T, van Heyningen V, Hastie ND, Meijers-Heijboer H, Drechsler M, Royer-Pokora B, Collins F, Swaroop A, Strong LC, Saunders GF (1991) Positional cloning and characterization of a paired box- and homeobox-containing gene from aniridia region. Cell 67:1059–1074

Wang RW, Newton DJ, Huskey S-EW, McKeever BM, Pickett CB, Lu AYH (1992) Site-directed mutagenesis of glutathione S-transferase YaYa. Important roles of tyrosine 9 and aspartic acid 101 in catalysis. J Biol Chem 267:19866–19871

Wilce MCJ, Parker MW (1994) Structure and function of glutathione S-transferases. Biochim Biophys Acta 1205:1–18

Wistow G, Piatigorsky J (1987) Recruitment of enzymes as lens structural proteins. Science 236:1554–1556

Wistow G, Piatigorsky J (1988) Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. Annu Rev Biochem 57:479–504

Wistow G (1993) Lens crystallin: gene recruitment and evolutionary dynamism. Trends Biochem Sci 18:301–306

Zuckerkandl E (1994) Molecular pathways to parallel evolution: I. Gene nexuses and their morphological correlates. J Mol Evol 39: 661–678

Zuker CS (1994) On the evolution of eyes: would you like it simple or compound? Science 265:742–743