Octopus S-crystallins with endogenous glutathione S-transferase (GST) activity: sequence comparison and evolutionary relationships with authentic GST enzymes

Shyh-Horng CHIOU,*†§ Chih-Wen YU,† Chou-Wen LIN,† Fu-Ming PAN,† Shao-Fan LU,* Hwei-Jen LEE‡ and Gu-Gang CHANG‡

*Laboratory of Crystallin Research, Institute of Biochemical Sciences, National Taiwan University, †Institute of Biological Chemistry, Academia Sinica, and

‡Department of Biochemistry, National Defense Medical Center, Taipei, Taiwan

S-Crystallin is a major protein present in the lenses of cephalopods (octopus and squid). To facilitate the cloning of this crystallin gene, cDNA was constructed from the poly(A)⁺ mRNA of octopus lenses, and amplified by PCR for nucleotide sequencing. Sequencing of 10 of 15 positive clones coding for this crystallin revealed three distinct S-crystallin isoforms with 61–64% identity in nucleotide sequences and 42–58% similarity in amino acid sequences when compared with homologous crystallins in squid lenses. These charge-isomeric crystallins also show between 26 and 33% amino acid sequence identity to four major classes of glutathione S-transferase (GST), a major detoxification enzyme present in most mammalian tissues. For further analysis,

expression of one of the S-crystallin cDNAs was carried out in the bacterial expression system pQE-30, and the S-crystallin protein produced in *Escherichia coli* was purified to homogeneity to determine the enzymic properties. We found that the expressed octopus S-crystallin possessed much lower GST activity than the authentic GSTs from other tissues. Sequence comparison and construction of phylogenetic trees for S-crystallins from squid and octopus lenses and various classes of GSTs revealed that S-crystallins represent a multigene family which is structurally related to Alpha-class GSTs and probably derived from the ancestral GST by gene duplication and subsequent multiple mutational substitutions.

INTRODUCTION

Glutathione S-transferases (GSTs; EC 2.5.1.18) are a group of ubiquitous proteins found in various animal and plant tissues which act as detoxification enzymes by catalysing the conjugation of glutathione with a large number of electrophilic alkylating compounds, thereby protecting cells against potential toxicity [1,2]. They represent products of several major classes of genes, i.e. Alpha, Mu, Pi and Theta. These GST isoenzymes are grouped with respect to their sequences, immunological properties and substrate specificities [3–5].

More than 90% of the proteins in the lenses of various animal species consist of a group of proteins called crystallins; these are water-soluble proteins and presumably have a structural role in the cytoplasm of lens fibre cells for the maintenance of transparency and optical clarity. Many crystallins have been identified, and some of them are expressed as complex isoforms with regard to either their size or charge heterogeneity. Three major families of crystallins, i.e. those classified as α -, β -, and γ -crystallins, are found in all vertebrates [6–8], and other taxon-specific crystallins with enzymic functions, e.g. δ -, ϵ -, λ - and τ -crystallins, have been found to be present in specific species of various classes of animals [9–11]. More intriguing is the finding that in the lenses of the invertebrate cephalopods (squid and octopus), the eyes of which are considered to be as complex as those of vertebrates, only two types of crystallins, designated Ω - [12] and S-crystallins [13–15], have been found. The former occurs only in octopus, and the latter in both octopus and squid. Despite the conventional view of the structural role of these lens crystallins, studies on the

chemistry of crystallins from various animal eye lenses seem to indicate that these structural proteins may be descended from enzymes of major metabolic or detoxification pathways [16]. For example, S- and Ω -crystallins are structurally similar to GST and aldehyde dehydrogenase respectively [17–19].

The present study aimed to make a systematic characterization of octopus S-crystallin using PCR methodology. This should aid in the structural analysis of multiple isoforms of S-crystallin from an invertebrate species that has the advanced camera-type image-forming lens also found in higher vertebrates. We have sequenced and expressed clones encoding major S-crystallins of octopus lenses having unusually high methionine contents (> 10%). Structural comparison of this presumably structural lens protein with GSTs from various sources and S-crystallins from another species of cephalopod (squid) revealed that the basic structural motif of GST is maintained to some extent in octopus and squid S-crystallins, despite the fact that these crystallins possess much lower enzyme activity than authentic GSTs.

MATERIALS AND METHODS

Isolation and characterization of octopus S-crystallin

Common octopuses (*Octopus vulgaris*) were obtained from local fishing ships. The lenses were dissected from the fresh octopuses and immediately stored in liquid nitrogen before processing for protein and mRNA isolation. The pooled lenses were de-capsulated and homogenized in 10–20 ml of 0.05 M Tris/sodium bisulphite buffer, pH 7.5, containing 5 mM EDTA as described

Abbreviations used: GST, glutathione S-transferase; CDNB, 1-chloro-2,4-dinitrobenzene; IPTG, isopropyl β -D-thiogalactopyranoside; OctS1, OctS2, OctS3, three isoforms of octopus S-crystallin.

§ To whom correspondence should be addressed: Institute of Biological Chemistry, Academia Sinica, P.O. Box 23-106, Taipei, Taiwan.

The nucleotide sequence data of the S-crystallin genes reported in this paper have been deposited in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession numbers X65543 (OctS1), X65544 (OctS2) and X74858 (OctS3).

previously [12]. The supernatant from centrifugation at 27000 *g* was adjusted to give a protein concentration of about 20–30 mg/ml, and a 5.0 ml aliquot was applied to a Fractogel TSK HW-55 column (Superfine Grade; Merck).

SDS/PAGE (5% stacking gel/14% resolving gel) was carried out as described by Laemmli [20] with some modifications. Isoelectric focusing was run on the Phast System using a Pharmacia dry IEF PhastGel (Pharmacia, Uppsala, Sweden) by mixing the gel with 1 ml of distilled water and 150 μ l of glycerol containing 1.5 mg of arginine and 80 μ l of Phamalytes of pH 8–10.5 and pH 3–10 (5:3, v/v) at 2.5 mA and 1300 V for 20 min.

Amino acid compositions were determined using a Beckman High-Performance Amino Acid Analyzer (Model 6300) with a dual-channel data system using a single column based on ion-exchange chromatography. The dialysed and lyophilized protein samples were hydrolysed at 160 °C for 45 min on a dry heating block using 4 M methanesulphonic acid containing 3-(2-amino-ethyl)indole (Pierce) for complete amino acid analysis from a single protein hydrolysate [21].

The N-terminal sequences of crystallin fractions from the gel-permeation column were determined by automated Edman degradation with a pulsed-liquid phase sequencer (Model 477A; Applied Biosystems). The lyophilized crystallin samples, each containing about 1–5 nmol of protein, were dissolved in 200 μ l of 0.1% trifluoroacetic acid or 0.1% SDS/0.1% trifluoroacetic acid (1:1, v/v), and 10 μ l samples were taken for sequence determinations.

Preparation of mRNA from lenses and cloning by PCR amplification

Two deep-frozen lenses from one octopus were homogenized and RNA was extracted according to standard procedures [22]. To obtain a full-length octopus S-crystallin gene, poly(A)⁺ mRNA was purified using a QuickPrep mRNA preparation kit (Pharmacia) and then a cDNA mixture was constructed using a RiboClone cDNA synthesis system (Promega, Madison, WI, U.S.A.). Two oligonucleotide primers with opposing orientations (covering the 5' and 3' coding regions of mature octopus S-crystallin partial sequences) [12,16,17,23], i.e. the forward primer 5'-CAATATCAAGCATGCCTTC(T/C)TACAC-3' and the reverse primer 5'-TTGTGTCTAGA(T/C)(T/C)TA(A/G)-AA(A/T/G)TC(A/G)GT(A/G)(T/C)-3', containing *Sph*I and *Xba*I restriction enzyme sites (underlined) respectively, were obtained from a commercial primer vendor. PCR reactions were carried out in a 100 μ l volume containing 0.1 μ g of cDNA template, 1 μ g of each primer, 0.2 mM of each dNTP, 2.5 units of *Taq* polymerase and other buffer components as recommended by Promega. The reactions were subjected to 40 cycles of heat denaturation at 94 °C for 1.5 min, annealing of the primers to the DNA at 50 °C for 1 min, DNA chain extension with *Taq* polymerase at 72 °C for 3 min, followed by a final extension at 72 °C for 10 min. The PCR products were separated on a 1.2% agarose gel, electroeluted and digested with *Sph*I and *Xba*I. The DNA fragment was subcloned into pUC19 previously digested with these two enzymes, and then transformed into JM109 of *E. coli*. Fifteen positive clones were selected and ten of them were sequenced by the dideoxynucleotide chain-termination method [24] using T₇ DNA polymerase (Sequenase Version 2.0; USB Corp., Cleveland, OH, U.S.A.).

The DNA sequences determined by a conventional isotope-labelling manual method were double-checked by automatic fluorescence-based sequencing of templates amplified by PCR using a model 373A DNA Sequencing System (Applied Bio-

systems) with a *Taq* DyeDeoxy terminator cycle sequencing kit (Applied Biosystems).

Cloning and expression of histidine-tagged octopus S-crystallin

A nucleotide segment coding for one of the octopus S-crystallin isoforms (OctS3) was synthesized by amplifying the S-crystallin open reading frame of the plasmid by using two primers designed for introducing cleavage sites for *Bam*HI and *Hind*III plus a 6 $\times$ His tag and an N-terminal extension of pentapeptide (Asp)₄-Lys. This PCR-synthesized nucleotide segment was cloned in-frame into the *Bam*HI/*Hind*III sites of the isopropyl β -D-thiogalactoside (IPTG)-inducible *E. coli* expression vector pQE-30. The resulting plasmid, pQE-30/S-crystallin, was transformed into *E. coli* strain M15 containing the pREP4 repressor plasmid. Transformants were selected on LB-agar plates supplemented with 25 μ g/ml kanamycin and 100 μ g/ml ampicillin. Recombinant plasmids were identified by restriction endonuclease analysis and clones containing the correctly sized insert were sequenced. For induction of gene expression, *E. coli* M15[pREP4] cells containing plasmid pQE-30/S-crystallin were grown at 37 °C in 1 litre of 2 $\times$ YT medium (yeast/tryptone culture medium; Difco) containing 25 μ g/ml kanamycin and 100 μ g/ml ampicillin. After reaching an *A*₆₀₀ of 0.5 the culture was divided into 4 $\times$ 250 ml portions and IPTG was added to a final concentration of 2 mM. The cultures were induced for a period of up to 2.5 h.

Purification of histidine-tagged octopus S-crystallin

Culture containing induced cells (500 ml) was centrifuged and the pellet was resuspended (5 ml/g) in buffer A (6 M guanidinium chloride, 0.1 M sodium phosphate, 0.01 M Tris/HCl, pH 8.0). The cells were stirred for 1 h at room temperature. This suspension was centrifuged at 10000 *g* for 15 min at 4 °C. An 8 ml portion of a slurry of nitrilotriacetic acid (Ni-NTA) resin (50%, w/v) was added to the supernatant, followed by stirring at room temperature for 45 min; the lysate was then loaded on to a 4 ml Ni-NTA column pre-equilibrated with buffer A and washed with buffer B (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris/HCl, pH 8.0) followed by buffer C (as for buffer B, except at pH 6.3). The bound material was eluted by buffer D (6 M guanidinium chloride, 0.2 M acetic acid). The chromatogram was monitored by UV absorbance at 280 nm and collected fractions were analysed by SDS/PAGE and Western blot analysis.

Immunoblotting analysis of expressed S-crystallin

Each lysed sample (100 μ l) was mixed with an equal volume of 10% (w/v) trichloroacetic acid, left on ice for 20 min and spun for 15 min in a Microfuge; the pellet was washed with 100 μ l of ice-cold ethanol, dried and resuspended in sample buffer. The samples were loaded immediately after incubation for 7 min at 95 °C. For immunodetection of S-crystallin, the gel was subjected to electroblotting on to a nitrocellulose membrane after SDS/PAGE followed by Coomassie Blue staining or immunological analysis using a rabbit antiserum against purified native S-crystallin obtained from a gel-permeation column and staining with horseradish peroxidase-conjugated goat anti-rabbit IgG (Sigma, St. Louis, MO, U.S.A.).

Sequence comparison of cephalopod S-crystallin and various GSTs

In the comparison and analysis of the deduced amino acid sequences from determined cDNA sequences coding for octopus

S-crystallins, a LaserGene software package (for the Apple Macintosh computer; from DNASTAR, Madison, WI, U.S.A.) was used for the estimation of homology based on percentage sequence similarity and divergence among nucleotide and protein sequences. Percentage divergence is calculated by comparing sequence pairs in relation to the phylogenetic tree. On the other hand percentage similarity is estimated by comparing sequences directly without taking account of phylogenetic relationships. Phylogenetic or evolutionary trees were constructed using the algorithm of Jotun Hein (Institute of Genetics and Ecology, University of Aarhus, Denmark) in the MegAlign programs of the package. This is a multiple sequence alignment program that builds trees as it aligns DNA or protein sequences using a combination of distance matrix and approximate parsimony methods. This method constructs multiple alignments by imposing restrictions based on the evolutionary relatedness of the aligned sequences, which is useful for the alignment of highly evolved gene families that have clear evolutionary relationships.

GST activity assay for octopus S-crystallins

GSH (Sigma) and 1-chloro-2,4-dinitrobenzene (CDNB; Serva) were purchased from the designated sources. S-Crystallin from octopus was purified to apparent homogeneity by a procedure similar to that used for squid S-crystallin [15]. The assay of the GST activity of native and expressed octopus S-crystallins was performed using a Hitachi spectrophotometer at 25 °C according to the established method [25].

RESULTS AND DISCUSSION

The discovery of crystallins with enzymic activity in various taxons of disparate species is part of the allure of crystallin research. Of vital importance, and no less interesting in a structural analysis of unique crystallins with some conventional or exotic enzymic functions, is the knowledge of their primary structures achieved by modern molecular cloning and PCR techniques. Only after such information is obtained will it be possible to shed some light on the specific functional role contributed by each individual crystallin in the lens. Several crystallins have been characterized both structurally and enzymically and shown to be related to conventional metabolic enzymes. Especially noteworthy are the ϵ - and δ -crystallins isolated from avian and reptilian lenses, which possess genuine catalytic activity and the defined kinetic mechanisms of lactate dehydrogenase and argininosuccinate lyase respectively [26,27]. In the present study we characterize and compare the S-crystallins of cephalopods with the GSTs of vertebrates; we have been pursuing the crystallin structures by conventional protein sequencing for many years [13,15], with some difficulty.

Isolation and characterization of octopus S-crystallin

Figure 1 shows the elution pattern of lens extracts from *Octopus vulgaris* separated on a TSK gel-permeation column. Three peaks were obtained for this octopus species, in contrast to only one major peak for squid lenses [23]. The native molecular masses of these fractions were determined to be 190 and 60 kDa for peaks 1 and 2 respectively; peak 3 was shown to be a mixture of non-protein small molecules. Peak 1 has been named Ω -crystallin and is reported to represent a tetrameric protein with four identical subunits of 59 kDa [12], which was recently shown to possess a primary structure similar to that of aldehyde dehydrogenase [19]. Peak 2, previously named S-crystallin due to it being the only major squid lens protein, was shown to be a dimeric protein with two subunits of about 25–29 kDa, similar to

Figure 1 Isolation and fractionation of octopus lens crystallins by gel-permeation chromatography

A Fractogel TSK HW-55(S) column (2.5 cm $\times$ 115 cm) was used. The elution buffer was 0.05 M Tris/sodium bisulphite buffer, pH 7.5, containing 5 mM EDTA as described previously [12]. The column eluates (3.2 ml/tube in 4.1 min) were monitored for absorbance at 280 nm. Three crystallin fractions (1–3) labelled correspond to Ω - and S-crystallins and a non-protein UV-absorbing substance respectively.

the squid crystallin characterized previously [13–15]. The apparent charge heterogeneity of purified S-crystallin from the rechromatographed peak 2 was revealed by high-resolution isoelectric focusing gel analysis (Figure 2). This indicated a multitude of charge-isomeric forms of at least 10 protein species, with a pI range of 5.8–9.8. Similarly, N-terminal sequence analysis of the first 25 amino acids of this rechromatographed S-crystallin fraction consistently indicated several heterogeneous sites with more than two amino acids being identified along the polypeptide

Figure 2 Isoelectric-focusing gel analysis of the S-crystallin fraction under native conditions

About 2 μ g each of S-crystallin (rechromatographed peak 2 in Figure 1) and pI protein markers (lane 2) was layered on a Pharmacia dry IEF PhastGel by mixing the gel with 1 ml of distilled water and 150 μ l of glycerol containing 1.5 mg of arginine and 80 μ l of Pharmalytes of pH 8–10.5 and pH 3–10 (5:3, v/v) under the conditions of 2.5 mA and 1300 V for 20 min (basic end is at the top). The pI of the S-crystallin fraction was estimated using a pI marker calibration kit (pI 3.5–9.3) as shown on the right. Arrows indicate stained bands corresponding to the precipitated S-crystallin (lane 1) and the dye front (lane 2).

Table 1 Comparison of amino acid compositions and pIs of native S-crystallin and its isoforms as determined from the cDNA sequences of crystallins from octopus and squid lenses

Data for native S-crystallin (peak 2 of Figure 1) are expressed as the number of residues per molecule of protein based on the presence of 10 Gly residues per molecule and the relative mol% as determined from the amino acid analysis (mean of duplicate determinations). Tryptophan content was determined by hydrolysis in 4 M methanesulphonic acid containing 3-(2-aminoethyl)indole. Amino acid composition data and pI values for the other crystallins were calculated from their protein sequences, deduced from their respective cDNA sequences. SL11 is an S-crystallin from squid lenses.

Amino acid	OctS1	OctS2	OctS3	SL11	S-crystallin
$\frac{1}{2}$ Cys	8	7	7	6	7
Asx	24	28	27	21	26
Thr	7	8	8	11	8
Ser	21	17	17	5	18
Glx	22	18	19	23	19
Pro	7	8	8	8	8
Gly	10	10	10	8	10
Ala	12	11	11	21	11
Val	6	4	4	7	4
Met	24	22	22	10	21
Ile	4	7	7	10	6
Leu	13	16	16	20	16
Tyr	11	11	11	11	10
Phe	11	12	12	11	12
His	6	6	6	3	6
Lys	4	4	3	16	3
Arg	23	24	25	11	23
Trp	0	1	1	2	1
Total	213	214	214	204	209
pI	9.205	8.905	8.951	9.015	5.8–9.8

chain (residues 15–25). Because of the complexity of these multiple charge-isomeric forms of S-crystallin, it is very difficult to obtain complete and unambiguous sequence assignment by conventional protein sequencing. Therefore we have determined the sequences of these multiple isoforms by the powerful techniques of PCR amplification and sequencing.

cDNA amplification by PCR and sequence analysis

PCR amplification of an octopus lens cDNA mixture with the designed primers based on the 5' and 3' coding regions of mature octopus S-crystallin partial sequences (covering 5' and 3' coding regions for partial sequences of the first three and last three amino acids in mature octopus S-crystallin) [12,16,17] achieved the isolation of one PCR fragment of about 680 bp (results not shown), corresponding to a complete reading frame encoding octopus S-crystallin of 210–220 amino acid residues. It is noteworthy that more than 10 positive clones have been identified, with their 5' coding nucleotide sequences being determined to be essentially identical to that of the designed 5' primer. We have sequenced 10 out of 15 clones and found only three distinct cDNA sequences encoding protein sequences of the required size (214–215 amino acids). Some of the main structural features of these deduced S-crystallin protein sequences with regard to GSTs of various classes will be discussed later. These three predominant S-crystallin isoforms (with pIs of 9.21, 8.91 and 8.95 for OctS1, OctS2 and OctS3 respectively) indeed possess N-terminal sequences (the first 25 N-terminal residues) closely similar to that determined for native S-crystallins (with heterogeneity sites detected for residues 15–25) isolated from lens homogenates, which should be representative of several distinct crystallin bands shown in Figure 2. It is worth noting that OctS2 and

OctS3 possess almost identical amino acid sequences (99.1% identity), whereas OctS1 exhibits a more diverse sequence (65.8% identity). Therefore there are probably two types of S-crystallin present in octopus lenses. We have calculated the amino acid compositions of OctS1, OctS2 and OctS3 plus one S-crystallin from squid species [28] and compared these compositions with that of native S-crystallin isolated from the octopus lens (Table 1). This clearly indicates a closer similarity in amino acid composition between the three isoforms of octopus S-crystallin compared with that of the squid protein, and native S-crystallin isolated from octopus lenses possesses a composition closer to that of OctS2 and OctS3 than OctS1.

Expression and immunological detection of cloned and native S-crystallins

Since there has so far been no report on the successful expression of recombinant S-crystallins from octopus or squid lenses, we have adopted a different approach by cloning the PCR-amplified fragment encoding the enzyme and expressing it in an *E. coli* expression vector. In order to check the identity of the expressed product, we analysed the antigenic structures of cloned S-crystallin by using the antiserum against purified S-crystallin from octopus lens (Figure 3). It is noteworthy that the antiserum against native S-crystallin cross-reacted specifically only with expressed products isolated at different times, and no reactions were detected with control lysate without IPTG induction. The molecular size of the expressed S-crystallin was found to be close to that of the native crystallin (25–27 kDa on SDS/PAGE) even if the six-histidine tag and the N-terminal extension pentapeptide (Asp)₄-Lys were not removed. However, the attachment of the histidine tag and the extension pentapeptide seems to account for part of the loss of enzymic activity in the expressed protein as compared with the native S-crystallin, despite their immunological close-relatedness.

By enzymic assay using CDNB as the electrophilic substrate, it was shown that the GST activity of the expressed protein was about one-tenth that of native S-crystallin, which is in turn only about one-twentieth of the GST activity of the total lens homogenate. Expression of eukaryotic proteins in *E. coli* often results in the formation of insoluble inclusion bodies, which makes the regeneration of enzymic activity difficult. This is especially true for S-crystallins, which possess seven or eight cysteine residues in each subunit; they became turbid and insoluble upon oxidation at ambient temperature [12,15]. We are currently refining the refolding conditions in order to isolate the active enzyme from the inclusion body of expressed S-crystallin dissolved in urea solution. The specific GST activity of purified S-crystallin was found to be about 0.10 $\mu\text{mol}/\text{min}$ per μg , compared with 100–200 $\mu\text{mol}/\text{min}$ per μg for most mammalian GST enzymes. However, the protein concentration of S-crystallin in the lens is tens of thousands fold greater than that of GSTs in various mammalian tissues. Therefore the limited GST activity *in vitro* shown here may have some biological significance in acting as a protective detoxification system for the lens *in vivo*, besides the structural role played by other enzyme crystallins simply because of their stability, as suggested previously [29,30].

Structural comparison of S-crystallins from octopus and squid lenses

The DNA sequences of the three S-crystallin isoforms (OctS1, OctS2 and OctS3) comprise the full-length coding regions encoding 214–215 amino acids. Among the 10 clones that we sequenced, the ratio of these three forms was 3:3:4, which indicated approximately equal abundance of these isoforms in

Figure 3 Expression and immunoblotting analysis of the histidine-linked S-crystallin product

The gels were subjected to electroblotting to a nitrocellulose membrane after SDS/PAGE followed by immunological analysis using rabbit antiserum against native S-crystallin isolated from peak 2 of Figure 1. Extracts were obtained from bacterial cultures containing the expression vector pQE-30/OctS3. Lane S, standard protein markers in kDa: phosphorylase *b* (94), BSA (66), ovalbumin (45), carbonic anhydrase (30), soybean trypsin inhibitor (20) and lysozyme (14). Lane 1, native S-crystallin; lane 2, expressed OctS3 crystallin from culture lysates 3 h after IPTG induction. (a) Coomassie Blue-stained gel; (b) Western immunoblot of SDS/PAGE gel with antiserum against native S-crystallin, stained with horseradish peroxidase-conjugated goat anti-rabbit IgG. Control culture lysate without IPTG induction showed no bands after immunoblotting.

		Similarity (%)									
Divergence (%)		1	2	3	4	5	6	7			
	1		80.0	99.0	90.0	80.0	61.0	64.0	1	OctS3	
	2	18.8		80.0	79.0	81.0	62.0	63.0	2	OctS1	
	3	2.0	20.0		90.0	80.0	61.0	64.0	3	OctS2	
	4	7.2	19.9	8.5		80.0	61.0	64.0	4	OL1	
	5	16.3	17.5	17.7	16.4		61.0	63.0	5	OL2	
	6	41.9	42.9	42.4	42.0	41.3		64.0	6	SL11	
	7	37.5	39.2	38.4	37.6	37.4	44.5		7	SL20	
		1	2	3	4	5	6	7			
		1	2	3	4	5	6	7			
Divergence (%)		1	2	3	4	5	6	7			
	1		67.0	99.0	91.0	80.0	53.0	58.0	1	OctS3	
	2	31.6		68.0	65.0	66.0	42.0	48.0	2	OctS1	
	3	0.3	31.4		92.0	80.0	54.0	58.0	3	OctS2	
	4	4.0	34.2	3.7		78.0	56.0	57.0	4	OL1	
	5	12.1	30.1	12.1	13.5		55.0	58.0	5	OL2	
	6	36.2	53.0	36.0	36.3	36.2		53.0	6	SL11	
	7	37.8	50.0	37.5	39.2	39.6	39.8		7	SL20	
		1	2	3	4	5	6	7			

Figure 4 Pair-wise comparison of the percentage similarity and divergence for nucleotide (a) and amino acid (b) sequences of S-crystallins using the multiple-sequence alignment program LaserGene

The percentage similarity between each sequence pair is read from the upper right area, and the percentage divergence is shown in the lower left area of the matrix table. OL1 and OL2 are S-crystallins from *Octopus dofleini* [16], and SL11 and SL20 are squid S-crystallins [28].

the expressed S-crystallins of octopus lens. However, OctS1 showed only about 80 % and 68 % identity in nucleotide and deduced amino acid sequences respectively with OctS2 or OctS3 (which were almost identical in nucleotide sequences). Figure 4 shows a pair-wise comparison of DNA (Figure 4a) and amino acid (Figure 4b) sequences with regard to the percentage sequence similarity and divergence for seven available S-crystallin sequences (five from octopus and two from squid) using a sequence analysis software package (LaserGene; DNASTAR). It is noteworthy that the extent of sequence similarity is greater between the octopus S-crystallins (67–99 %) than between octopus and squid proteins (42–58 %) (Figure 4b). When comparing these three crystallin sequences with two S-crystallin sequences (OL1 and OL2) from another octopus species, *Octopus dofleini*, reported by Tomarev et al. [16], the similarity of amino acid sequences was 65–92 %, revealing that S-crystallins from different octopus species exhibit a high degree of heterogeneity, similar to that found for squid S-crystallins and various GSTs from different sources.

In general, nucleotide pairs show a higher percentage sequence similarity than their deduced amino acid sequence pairs (Figures 4a and 4b), probably due to the degeneracy in the triplet codes for various amino acids. Therefore detection of sequence divergence based on protein sequences rather than cDNA sequences seems to be more sensitive when comparing highly homologous protein families. We have constructed phylogenetic trees for these homologous S-crystallins from squid and octopus based on the percentage divergence between the DNA and protein sequences (Figures 5a and 5b). The results indicate that both trees indeed exemplify the close relatedness between S-crystallin pairs from either the octopus or the squid; there is a more remote relationship between octopus and squid S-crystallins. It is of interest to note that OctS1 is located in a branch position away from OctS2 and OctS3, since it only shows about 80 % nucleotide and 67–68 % amino acid sequence similarity to these two crystallins. There is a slight discrepancy between the trees constructed based on DNA (Figure 5a) and amino acid (Figure 5b) sequences, especially with regard to the relative positions of OctS1 and OL2 in the trees. It is concluded that OctS1 probably diverged from the ancestral S-crystallin gene long before the evolution of the OctS2/OctS3 or the OL1/OL2 pair. It is also clearly demonstrated that the squid S-crystallins SL20 and SL11 indeed followed a different evolutionary path from that of octopus crystallins. We would like to emphasize that these phylogenetic trees were constructed by using the algorithm of Jotun Hein, which is a multiple sequence alignment program that builds trees as it aligns DNA or protein sequences using a combination of distance matrix and approximate parsimony methods. This method constructs multiple alignments by imposing restrictions based on the evolutionary relatedness of the aligned sequences, which is useful in the alignment of highly evolved gene families that have clear evolutionary relationships as in the case of the S-crystallin family.

Structural comparison of S-crystallins and various classes of GSTs

All three octopus crystallins (214–215 amino acids; 24.6 kDa) reported here and squid S-crystallins (204–222 amino acids) [28] show some 25–30 % sequence similarity to the rat GST Yb subunit (218 amino acids) [31] in a pair-wise comparison of their amino acid sequences [17]. S-Crystallins from octopus and squid lenses are composed of homo- or hetero-dimers, with subunit molecular masses of about 25–30 kDa [12,14,15]. Similarly, the GSTs are a multigene family of dimeric multifunctional proteins

Figure 5 Phylogenetic trees of seven S-crystallins from two octopus species and one squid species

These trees were constructed based on the percentage divergence between DNA (a) and protein (b) sequences using a combination of distance matrix and approximate parsimony methods. This technique carries out multiple alignment by imposing restrictions based on the evolutionary relatedness of the aligned sequences [39]. The length of each pair of branches represents the sequence distance between aligned pairs. The scale beneath the tree measures the evolutionary distance between sequences (in millions of years).

Figure 6 Multiple sequence alignment and comparison of seven S-crystallin sequences from two octopus species and one squid species

The identical amino acid residues among the various sequences, based on the first one (OctS3), are shown as white letters on a black background. The gaps were introduced for optimal alignment and maximum homology. Note that the sequences of OctS2 and OctS3 are very similar, but that of OctS1 is less so.

of 25–27 kDa found in all organisms, and they are believed to play a major role in the detoxification of xenobiotics [2,32,33]. GSTs comprise various isoenzymes containing several similar subunits in different combinations [34]. It is worth noting that the three octopus S-crystallins show 20–33% similarity in amino acid sequence and 32–44% similarity in nucleotide sequence

when compared with mammalian GSTs, underlying the remote evolutionary relationship between these two types of proteins. At least four gene classes, i.e. Alpha, Pi, Mu and Theta, constitute the GST gene superfamily. The high multiplicity of the GST isoenzymes provides a system for an evolutionary comparison with S-crystallins from cephalopods, supposedly recently evolved

Figure 7 Multiple sequence alignment and sequence comparison of octopus S-crystallins and four major classes of mammalian GSTs

Identical amino acid residues among the various sequences, based on the first one (OctS3), are shown as white letters on a black background. The gaps were introduced for optimal alignment and maximum homology. Arrows and X signs indicate 26 invariant or conservative amino acid residues, which are likely to be important for the structure and function of authentic GSTs of these four major classes (Alpha, Pi, Mu and Theta) [38]. Of these 26 residues, 21 were found to be invariant or conservatively substituted (arrows) in OctS1–OctS3, while the remaining five (X) have been mutated or deleted during evolution.

GST-like proteins. We have attempted to align the sequences of S-crystallins and major classes of GSTs for optimal homology.

Figure 6 shows the optimal alignment along the entire lengths of each relevant S-crystallin sequence by introducing a minimum number of gaps. It is clear from Figure 6 that, with the exception of OctS1 and SL20 (which have long stretches of mismatches), the correspondence between each pair of sequences is generally high, regardless of their origins. Especially noteworthy is the high degree of sequence identity between the five octopus S-crystallin sequences, again underlining the close relatedness of these proteins. On the other hand, the blocks of sequence identity are more dispersed along the lengths of the protein sequences for each representative class of GSTs when comparing these enzymes with the three octopus S-crystallins reported here (Figure 7).

In Figure 7 we have aligned the sequences of octopus S-crystallins with those of the four major GST classes. The sequence identity between these two families of proteins is low, with 28 %, 21 %, 20 % and 7 % identity for the pairs OctS3/Alpha GST, OctS3/Pi GST, OctS3/Mu GST, and OctS3/Theta GST respectively. Therefore it appears that the octopus S-crystallin family is only remotely related to the authentic GST enzymes, probably diverging from the ancient Alpha-class GST through multiple mutational changes to the basic primary structure of the latter.

The crystal structures of GSTs from three of the four major classes (with the exception of the Theta class) have been

determined [35–38]. Several residues are conserved in all authentic GSTs from different sources; these are Tyr-7, Pro-53, Asp-57, Ile-68, Gly-145 and Asp-152 (based on the numbering system for the human placental Pi-class enzyme), with Tyr-7 and Pro-53 playing key roles in the GSH binding site. As shown in Figure 7, the octopus crystallins OctS2 and OctS3 (but not OctS1) have retained these structural residues for substrate binding and structural stability. In a more detailed analysis of the structures of GSTs by Dirr et al. [38], it was found that, despite extensive substitution between different classes of GSTs, the fundamental structural features necessary for preserving the GST structural fold are maintained by 26 invariant residues, which are most probably of structural and functional significance. Of these 26 residues (indicated by arrows and X signs in Figure 7), 21 residues were found to be invariant in the octopus crystallins, with five amino-acid residues substituted or mutated during the process of evolution. It would be of interest to carry out site-directed mutagenesis at these five positions to determine whether they are responsible for the GST enzymic activity found in these lens crystallins.

Conclusion and perspectives

In conclusion, we have established the nucleotide and amino acid sequences of three cDNA clones encoding the major S-crystallins in octopus lenses. These crystallins appear to be related to the important detoxification enzymes GSTs, both in structure and in

evolutionary relationships. Especially intriguing is the conservation of invariant amino acid residues required for structural and functional properties. A kinetic study of S-crystallins indeed detected low yet genuine GST activity for these presumably abundant structural proteins. Further studies on the expression of octopus S-crystallins in cell culture systems, coupled with site-directed mutagenesis of these PCR-amplified clones, may be conducive to unravelling the intriguing evolutionary process leading to diminution of enzymic activity and the recruitment of these dual-function crystallins in the lenses of invertebrate Cephalopods.

This work was supported in part by Academia Sinica and the National Science Council, Taipei, Taiwan (grants NSC 83-0418-B-001-020-BA, NSC 83-0203-B-001-086 and NSC 84-2311-B-001-050BA).

REFERENCES

- Mannervik, B. and Danielson, U. H. (1988) *CRC Crit. Rev. Biochem.* **23**, 283–337
- Pickett, C. B. and Lu, A. Y. H. (1989) *Annu. Rev. Biochem.* **58**, 743–764
- Mannervik, B. (1985) *Adv. Enzymol. Relat. Areas Mol. Biol.* **57**, 357–417
- Tsuchida, S. and Sato, K. (1992) *Crit. Rev. Biochem. Mol. Biol.* **27**, 337–384
- Rushmore, T. H. and Pickett, C. B. (1993) *J. Biol. Chem.* **268**, 11475–11478
- Harding, J. J. and Dilley, K. J. (1976) *Exp. Eye Res.* **22**, 1–73
- Bloemendal, H. (1982) *CRC Crit. Rev. Biochem.* **12**, 1–38
- Chiou, S.-H. (1986) *FEBS Lett.* **201**, 69–73
- Wistow, G. J., Mulders, J. W. M. and de Jong, W. W. (1987) *Nature (London)* **326**, 622–624
- Wistow, G. J. and Piatigorsky, J. (1988) *Annu. Rev. Biochem.* **57**, 479–504
- Bloemendal, H. and de Jong, W. W. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* **41**, 259–281
- Chiou, S.-H. (1988) *FEBS Lett.* **241**, 261–264
- Chiou, S.-H. and Bunn, H. F. (1981) *Invest. Ophthalmol. Vis. Sci. ARVO Suppl.* **20**, 138
- Siezen, R. J. and Shaw, D. C. (1982) *Biochim. Biophys. Acta* **704**, 304–320
- Chiou, S.-H. (1984) *J. Biochem. (Tokyo)* **95**, 75–82
- Tomarev, S. I., Zinovieva, R. D. and Piatigorsky, J. (1991) *J. Biol. Chem.* **266**, 24226–24231
- Lin, C. W. and Chiou, S. H. (1992) *Biochem. Int.* **27**, 173–178
- Tomarev, S. I., Zinovieva, R. D. and Piatigorsky, J. (1992) *J. Biol. Chem.* **267**, 8604–8612
- Zinovieva, R. D., Tomarev, S. I. and Piatigorsky, J. (1993) *J. Biol. Chem.* **268**, 11449–11455
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Chiou, S.-H. and Wang, K.-T. (1988) *J. Chromatogr.* **448**, 404–410
- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Chiou, S.-H., Chang, W.-C., Pan, F.-M., Chang, T. and Lo, T.-B. (1987) *J. Biochem. (Tokyo)* **101**, 751–759
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Habig, W. H. and Jakoby, W. B. (1981) *Methods Enzymol.* **77**, 398–405
- Chiou, S.-H., Lee, H.-J. and Chang, G.-G. (1990) *Biochem. J.* **267**, 51–58
- Lee, H.-J., Chiou, S.-H. and Chang, G.-G. (1992) *Biochem. J.* **283**, 597–603
- Tomarev, S. I. and Zinovieva, R. D. (1988) *Nature (London)* **336**, 86–88
- Piatigorsky, J. and Wistow, G. (1991) *Science* **252**, 1078–1079
- Piatigorsky, J. (1992) *J. Biol. Chem.* **267**, 4277–4280
- Abramovitz, M. and Listowsky, I. (1987) *J. Biol. Chem.* **262**, 7770–7773
- Waxman, D. J. (1990) *Cancer Res.* **50**, 6449–6454
- Armstrong, R. N. (1991) *Chem. Res. Toxicol.* **4**, 131–140
- Mannervik, B. and Jensson, H. (1982) *J. Biol. Chem.* **257**, 9909–9912
- Ji, X., Zhang, P., Armstrong, R. N. and Gilliland, G. L. (1992) *Biochemistry* **31**, 10169–10184
- Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G., Olin, B., Mannervik, B. and Jones, T. A. (1993) *J. Mol. Biol.* **232**, 192–212
- Wilce, M. C. J. and Parker, M. W. (1994) *Biochim. Biophys. Acta* **1205**, 1–18
- Dirr, H., Reinemer, P. and Huber, R. (1994) *Eur. J. Biochem.* **220**, 645–661
- Hein, J. (1990) *Methods Enzymol.* **183**, 626–645