Journal of Protein Chemistry, Vol. 7, No. 4, 1988

The Protein Sequence Homology of γ-Crystallins among Major Vertebrate Classes and Their DNA Sequence Homology to Heat-Shock Protein Genes

Shyh-Horng Chiou[1]

Received November 3, 1987

A systematic characterization of lens crystallins from five major classes of vertebrates was carried out by exclusion gel filtration, cation-exchange chromatography and N-terminal sequence determination. All crystallin fractions except that of γ-crystallin were found to be N-terminally blocked. γ-Crystallin is present in major classes of vertebrates except the bird, showing none, or decreased amounts, of this protein in chicken and duck lenses, respectively. N-Terminal sequence analysis of the purified γ-crystallin polypeptides showed extensive homology between different classes of vertebrates, supporting the close relatedness of this family of crystallin even from the evolutionarily distant species. Comparison of nucleotide sequences and their predicted amino acid sequences between γ-crystallins of carp and rat lenses and heat-shock proteins demonstrated partial sequence homology of the encoded polypeptides and striking homology at the gene level. The unexpected strong homology of complementary DNA (cDNA) lies in the regions coding for 40 N-terminal residues of carp γ-II, rat γ2-1, and the middle segments of 23,000- and 70,000-M_r heat-shock proteins. The optimal alignment of DNA sequences along these two segments shows about 50% homology. The percentage of protein sequence identity for the corresponding aligned segments is only 20%. The weak sequence homology at the protein level is also found between the invertebrate squid crystallin and rat γ-crystallin polypeptides. These results pointed to the possibility of unifying three major classes of vertebrate crystallins into one α/β/γ superfamily and corroborated the previous supposition that the existing crystallins in the animal kingdom are probably mutually interrelated, sharing a common ancestry.

KEY WORDS: γ-crystallin; heat-shock protein; sequence homology; multigene family; phylogenetic comparison; divergent evolution.

1. INTRODUCTION

Characterization of lens crystallins from the vertebrate and invertebrate lenses may provide a good testing ground of two major hypotheses of protein evolution (Chiou, 1986). The morphological similarity and immunological non-cross-reactivity of lens proteins from these two groups of lenses led to the prevalent view of convergent evolution of crystallins in the animal kingdom (Packard, 1972). Previous studies from this laboratory showed the relatedness of amino acid compositions between the squid and carp γ-crystallins, suggesting the possible sequence homology in the

[1] Institute of Biochemical Sciences, National Taiwan University, and Institute of Biological Chemistry, Academia Sinica, P.O. Box 23-106, Taipei, Taiwan, Republic of China.

0277-8033/88/0800-0527$06.00/0

primary structures of these crystallins. Pairwise statistical comparison of amino acid contents of various crystallins seemed in favor of divergent rather than convergent evolution, as commonly assumed in the literature (Siezen, 1981; Siezen and Shaw, 1982). However, the critical evidence in support of the divergent aspect of evolution still lies in finding the unequivocal sequence homology at the DNA or protein level between lens crystallins. In this paper, the N-terminal sequences of various γ-crystallins from different classes of vertebrates and the squid crystallin of invertebrate were determined; the segments of known partial sequences were then used for segmentwise comparison with the available sequences in the protein or gene databases. Despite the weak sequence homology found between the N-terminal segment of squid crystallin and those of vertebrate γ-crystallins, the match of several small segments is too close to be the result of fortuitous coincidence. Even more surprising is the presence of extensive homology in the nucleotide sequences between DNA sequences encoding N-terminal segments of various vertebrate γ-crystallins and the middle region of heat-shock protein gene (Ingolia and Craig, 1982).

2. MATERIALS AND METHODS

Squid (*Sepia esculenta*), carp (*Cyprinus carpio*), bullfrog (*Rana catesbeiana*), caiman (*Caiman crocodylus apaporiensis*), duck (mule duck, a hybrid between *Cairina moschata* and *Anas platyrhynchos var. domestica*), and cow (*Bos taurus*) lenses were obtained from local farms of animal husbandry or fish markets. The decapsulated lenses were homogenized in 10–20 ml 0.05 M Tris–Na bisulfite buffer, *p*H 7.5 containing 5 mM EDTA as described earlier by Chiou *et al.* (1979). The supernatant from 27,000 × *g* centrifugation was adjusted to give a concentration of about 20–30 mg/ml, and the 5.0 ml aliquot was applied to Fractogel TSK HW-55 (Superfine Grade, Merck). The γ-crystallin fraction from gel-filtration column was further separated into its subfractions on TSK CM-650 (S) cation-exchange column with a linear gradient of 0.05–0.25 M ammonium acetate in the presence of 0.1% 2-mercaptoethanol, *p*H 5.9.

Sodium dodecyl sulfate (SDS) polyacrylamide slab gel (5% stacking/14% resolving gel) was as described by Laemmli (1970), modified by Chiou (1987*a*). The amino acid compositions were determined with the LKB-4150 amino acid analyzer using a single-column system. The dialyzed and lyophilized protein samples were hydrolyzed at 150°C in evacuated tubes with constant-boiling 6 N HCl (Pierce Chemical Co.) for 1.5 hr (Chiou, 1988). Half-cystine was determined separately after performic acid oxidation. Tryptophan was determined with the procedure of Simpson *et al.* (1976) using 4 N methanesulfonic acid containing 0.2% tryptamine (Pierce Chemical Co.).

The N-terminal sequences of the major fractions from the gel-permeation column and the fractionated γ-crystallin subfractions from TSK CM-650 (S) cation-exchange chromatography were carried out by automated Edman degradation with a microsequencing sequenator (model 477A, Applied Biosystems). The lyophilized crystallin samples each containing about 1–5 nmoles protein were dissolved in 200 μl 0.1% trifluoroacetic acid (TFA) or 0.1% SDS/0.1% TFA (1 : 1 v/v) and 10 μl each for sequence determinations.

The computer-based sequence comparison of the 25-residue N-terminal segment of squid crystallin with those sequences in the protein database of the National Institute of Health (NIH) was carried out using the commercially available program. The nucleotide sequences of the reported rat γ2-1 crystallin gene (den Dunnen *et al.*, 1986) and carp γ-crystallin (in preparation) were also compared with the DNA sequences in the database of GenBank.

3. RESULTS AND DISCUSSION

Previous N-terminal sequence analyses of γ-crystallin subfractions from several different species demonstrated the extensive homology present in the amino-terminal segments of fish and mammalian γ-crystallins (Chiou *et al.*, 1986, 1987), suggesting a common ancestry for this class of vertebrate crystallins. The enlightening comparison of β- and γ-crystallins based on a combination of X-ray diffraction data and protein sequences has prompted the hypothesis that β- and γ-crystallins probably form a superfamily of β/γ-crystallins (Wistow *et al.*, 1981). In addition, Wistow *et al.* (1985) and Crabbe (1985) reported that sequence homologies has been found between a bacterial protein plus an oncogene protein and β/γ-family of crystallins, α-Crystallin was also shown to be homologous to small heat-shock proteins with regard to their sequences (Ingolia and Craig, 1982). However, there has been no success in correlating α- and β/γ-crystallins into a unified single superfamily of these three classes of crystallins.

Figure 1 shows typical elution patterns of lens extracts from four major classes of vertebrates and one class of invertebrates. Only one fraction was obtained for the squid lens of invertebrates, in contrast to three for the carp and five for the pig. The last peaks of the major eluted fractions for all vertebrates except the bird contain γ-crystallin of about 20,000 M_r, as judged from the subunit analysis by SDS-polyacrylamide gel electrophoresis (PAGE) (Chiou, 1986). The charge heterogeneity can be detected for γ-crystallins from all species, showing at least four charge-isomeric forms (Björk, 1964; Chiou *et al.*, 1986, 1987). Since γ-crystallin of the carp lens shows a striking similarity of amino acid compositions with the only crystallin found in the squid lens (Chiou, 1984), efforts have been made to correlate the primary structures of these two classes of crystallin.

It is prerequisite to obtain the detailed information in the protein or DNA sequences of different crystallins from various species in order to establish the phylogenetic relationship concerning the evolution of crystallins in general. Figure 2 showed the N-terminal sequences of γ-crystallins from various species encompassing all major classes of vertebrates. γ-Crystallin was found to be absent or present in much lesser amounts in the bird lenses and therefore was not included for comparison. It is evident that γ-crystallins obtained from the evolutionarily distant species still exhibit sequence homology in their N-terminal segments, indicating the close relatedness among various γ-crystallins of vertebrates.

Table I summarizes the amino acid compositions of heat-shock protein (Hsp 23), α-crystallin subunits (αA_2 and αB_2), carp γ-II, rat γ2-1, and squid crystallins. Previous statistical comparison of the reliability of phylogenetic information deduced from amino acid sequences and compositions (Cornish-Bowden, 1977, 1980) has

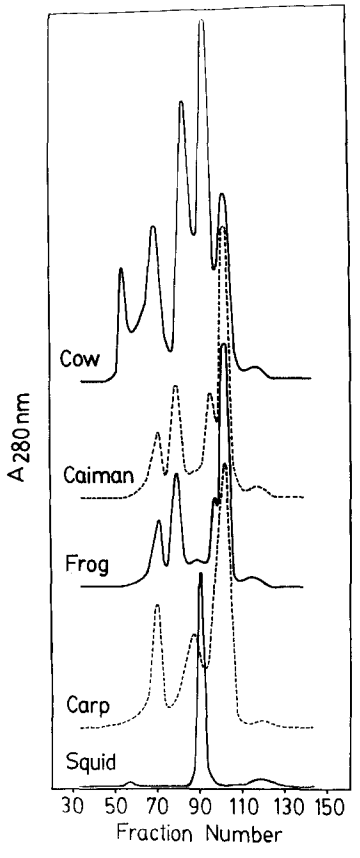

Fig. 1. Comparative gel-permeation chromatography on Fractogel TSK HW-55(S) (2.5 × 115 cm column) of lens extracts from the lenses of four species of vertebrates and one species of invertebrate. The column eluates (3.9 ml/tube per 4.6 min) were monitored for absorbance at 280 nm. The peak fractions eluted at about tube no. 103 for the four species of vertebrates represent γ-crystallins, the only major peak in the squid lens is the squid crystallin. These fractions were collected and used for amino acid analyses (see Table I). The absorbances at 280 nm are relative concentrations in arbitrary units. The small broad peaks after the γ-crystallin peak are nonprotein components of low-molecular masses.

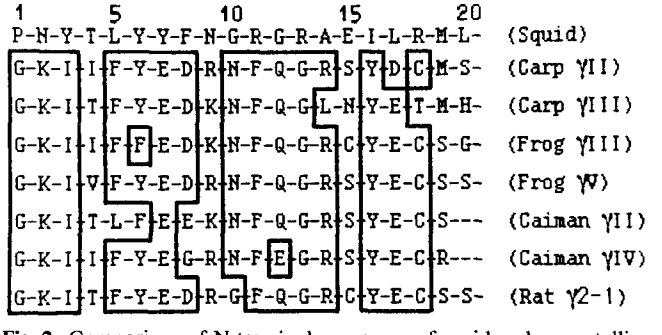

Fig. 2. Comparison of N-terminal sequences of squid and γ-crystallins from various species. The sequences listed were taken from this study and from previous reports (Chiou *et al.*, 1986, 1987), with the exception that the sequence of rat γ-crystallin was from den Dunnen *et al.* (1986). The sequences of the subfractions of γ-crystallins from each species were determined on the purified fractions upon cation-exchange chromatography of crude γ-crystallins from gel-permeation column. The homology regions are boxed. Amino acid residues are denoted by one-letter symbols.

Table I. Comparison of Amino Acid Composition[a]

Amino acids (mole %)	αA_2	Hsp23	αB_2	Rat γ2-1	Carp γ-II	Squid
1/2Cys	0.6	1.1	0	5.2	3.2	2.4
Asx	9.4	9.1	7.4	9.3	10.9	14.0
Thr	2.9	1.6	4.0	1.7	2.3	2.9
Ser	13.3	7.5	9.7	8.1	6.8	6.1
Glx	9.8	14.0	9.7	11.6	10.5	8.6
Pro	6.9	7.0	9.7	3.5	4.1	4.5
Gly	5.9	9.7	4.6	7.5	7.1	6.6
Ala	3.5	5.9	5.1	1.7	1.5	4.3
Val	5.7	9.7	5.7	5.2	4.1	1.7
Met	1.2	2.7	1.1	3.5	10.5	12.5
Ile	5.2	4.3	5.7	2.3	4.4	2.6
Leu	8.1	8.6	8.6	8.1	3.2	5.0
Tyr	3.4	3.2	1.1	8.7	6.8	6.5
Phe	8.1	2.7	7.4	3.5	5.9	6.2
His	4.1	2.7	5.1	4.1	3.4	1.8
Lys	4.1	5.4	5.7	2.3	2.5	5.3
Arg	7.5	4.8	8.0	11.6	10.9	7.9
Trp	0.6	0	1.1	2.3	2.1	1.4
$S\Delta Q$[b]	133	131	218	99	59	

[a] Data are taken from this study and from Bloemendal (1977), Ingolia and Craig (1982), den Dunnen *et al.* (1986), and the references cited therein.

[b] $S\Delta Q$ represents the pairwise comparison of amino acid contents of the adjacent crystallins as described by Chiou (1986).

indicated the useful applicability of the amino acid compositions in the phylogenetic study of different groups of proteins. A crude measure of the relatedness of different families of proteins can be exemplified by the values of $S\Delta Q$, i.e., sum of the squares of differences of each amino acid in the composition data (mole %) in the protein pair being compared (Marchalonis and Weltman, 1971). It is noteworthy that the similarity in amino acid content is evident between the pair with $S\Delta Q < 100$, e.g., $S\Delta Q = 41$ between αA_2 and αB_2 of α-crystallin and 59 for the carp γ-II/squid pair. It is of interest to note that the sequence comparison of the pairs of αA_2/Hsp23 and αB_2/Hsp23 with quite different amino acid compositions (an $S\Delta Q$ of 130) has shown more than 50% sequence similarity on a segment of 76 residues (Ingolia and Craig, 1982). This would indicate that the protein pairs being compared without apparent similarity in amino acid composition are sometimes related at the level of primary sequence. However, it is generally true that the protein pairs with small $S\Delta Q$ (<50) are somehow related at their primary sequences. Another prominent example is the comparison of βBp and calf γ-II ($S\Delta Q = 177$), which showed a high degree of sequence homology in their basic two-domain structures (Wistow *et al.*, 1981).

Since an $S\Delta Q$ of 177, similar to that of the closely related βB_p/calf γ-II pair, is obtained between Hsp23 and rat γ2-1, a detailed DNA sequence comparison of rat γ2-1 and the known sequences of the database in the GenBank has been carried out on a minicomputer-based comparison program. Figure 3 shows the optimal

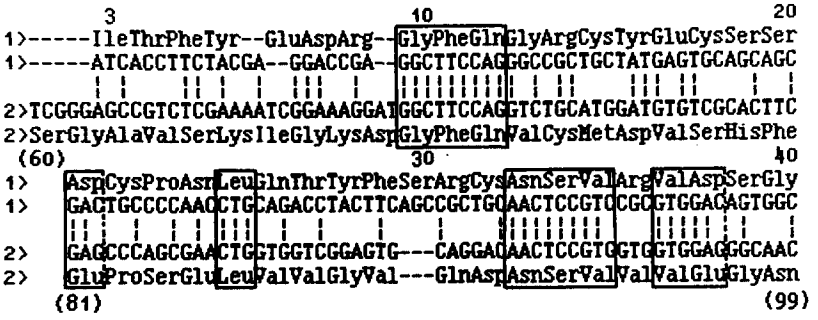

Fig. 3. Comparison of the nucleotide and amino acid sequences of rat γ2-1 crystallin (den Dunnen *et al.*, 1986) and heat-shock protein (Hsp23) of *Drosophila* (Ingolia and Craig, 1982). Sequences 1> and 2> represent the protein and DNA sequences of rat crystallin (3–40) and heat-shock protein (60–99), respectively. The vertical dashes indicate the DNA sequence identity of two genes. Identical and similar amino acid residues at the corresponding positions are boxed.

alignment of nucleotide sequences and their corresponding predicted amino acid sequences. The DNA sequence homology in this segment of 115 bases is about 50% (58 of 115) and protein sequence homology is 20% (8 of 40 amino acid residues). There are complete sequence identities at two noninterrupted segments of 10 and 9 nucleotides. They could not have arisen by fortuitous chance, as judged by the probability calculation of their concerted appearance in the DNA ($1/4^{10} \times 1/4^{9} = 1$ in $10^{11.4}$). Similarly, the protein sequence homology of 20% containing two peptide identities of three residues (boxed region of Fig. 3) are probably not coincidence in nature (the chance appearance of Gly-Phe-Gln and Asn-Ser-Val together at a randomized sequence will be $1/2^{14} \times 3/2^{14}$ or about 1 in 10^5, by taking into account of the codon frequency for each encoded amino acid). It would seem more impressive for the homology of DNA sequence as compared with that of protein sequence in this case. The extensive homology at the nucleotide level seems higher than usual, given the general degeneracy of some genetic codes. The homology observed for γ-crystallin and heat-shock protein could provide some insight into the elusive interrelationship between α and β/γ crystallins, as it had been reported that α-crystallin shares sequence homology with the heat-shock protein family (Ingolia and Craig, 1982; Russnak *et al.*, 1983).

Figure 4 lists the partial sequences of two carp γ-crystallin subfractions, rat γ2-1 and squid crystallins. It is clear that the γ-crystallin family of vertebrates shows strong homology in N-terminal segments, yet only weak homology is observed between carp and squid crystallin in this region. By contrast, we have detected the homology of the N-terminal region in the squid crystallin with the middle segment of rat γ-crystallin (boxed regions in Fig. 4). It is to be expected that more sequence homology will be found between the squid and carp γ-crystallins, if their detailed sequences are determined in the future. Recently, we sequenced one of the carp complementary DNA (cDNA) clones encoding the major γ-II crystallin subfraction with a high methionine content (in preparation). The derived amino acid sequence showed more than 50% homology with rat γ-crystallin; the corresponding cDNA

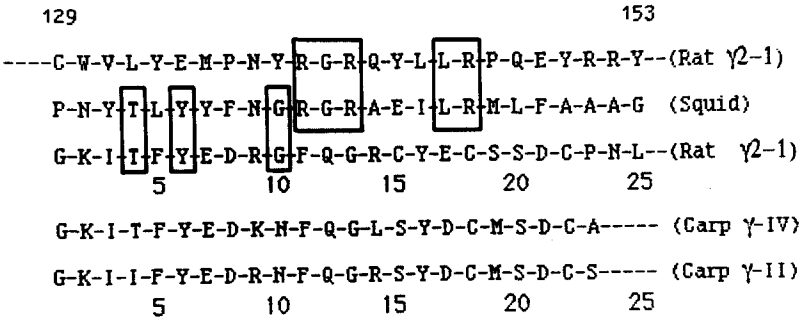

Fig. 4. Comparison of the amino acid sequences of squid, carp γ and rat γ2-1 crystallins. The sequences listed are the amino-terminal segments of these crystallins, except that the polypeptide fragment (amino acid residues 129–153) of rat γ2-1 is also listed above the squid sequence for comparison. The regions of sequence identity with optimal alignment are boxed. Amino acid residues are represented by one-letter symbols.

also exhibited the same type of segmentwise nucleotide homology with human heat-shock protein (Fig. 5). The percentage of protein sequence identity for the corresponding aligned segments is also found to be much lower (18%) than that of nucleotide sequences (48%).

In conclusion, the sequence comparison of γ-crystallins of vertebrates with heat-shock protein has demonstrated strong homology at the DNA sequences and partial homology of amino acid sequences. The weak homology of small sequence segments is also observed between the invertebrate squid crystallin and mammalian γ-crystallins. These results have provided a crucial link for the mutual relatedness of α, β/γ, and heat-shock proteins. The detailed protein and gene analyses of invertebrate squid crystallin should provide a firm basis in the construction of a unified phylogenetic tree and shed some light on the evolution of lens crystallins.

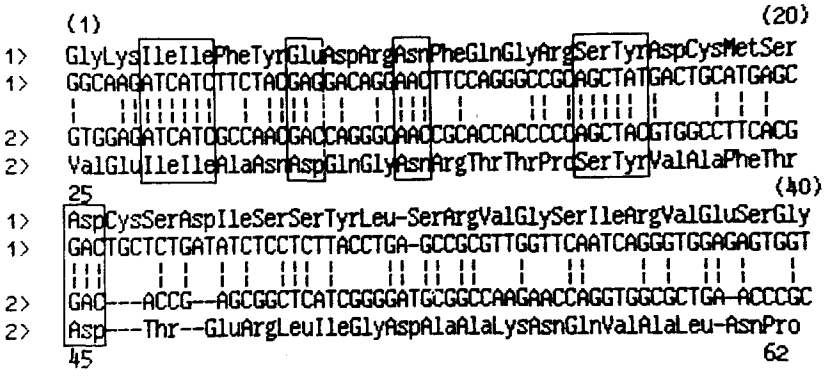

Fig. 5. Comparison of the nucleotide and amino acid sequences of carp γ-II crystallin and human heat-shock protein (Hsp70) (Hunt and Morimoto, 1985). Sequences 1> and 2> represent the protein and DNA sequences of carp crystallin (1–40) and heat-shock protein (25–62), respectively. The vertical dashes indicate the DNA sequence identity of two genes. Identical and similar amino acid residues at the corresponding positions are boxed.

ACKNOWLEDGMENTS

The author thanks Dr. M.-Z. Lai of the Department of Biology, Massachusetts Institute of Technology, for suggesting the sequence comparison of the N-terminal segment of squid crystallin with the known sequences in the available database. This work was supported in part by Academia Sinica and the National Science Council, Taipei, Taiwan, Republic of China.

REFERENCES

Björk, I. (1964). *Exp. Eye Res.* **3**, 254–261.
Bloemendal, H. (1977). *Science* **197**, 127–138.
Chiou, S.-H. (1984). *J. Biochem. (Tokyo)* **95**, 75–82.
Chiou, S.-H. (1986). *FEBS Lett.* **201**, 69–73.
Chiou, S.-H. (1987*a*). *Lens Res.* **4**, 163–171.
Chiou, S.-H. (1988). *Anal. Biochem.* in press.
Chiou, S.-H., Azari, P., Himmel, M. E., and Squire, P. G. (1979). *Int. J. Peptide Protein Res.* **13**, 409–417.
Chiou, S.-H., Chen, S.-W., and Lo, T.-B. (1986). *FEBS Lett.* **209**, 107–110.
Chiou, S.-H., Chang, W.-P., Lo, C.-H., and Chen, S.-W. (1987). *FEBS Lett.* **221**, 134–138.
Cornish-Bowden, A. (1977). *J. Theoret. Biol.* **65**, 735–742.
Cornish-Bowden, A. (1980). *Biochem. J.* **191**, 349–354.
Crabbe, M. J. C. (1985). *FEBS Lett.* **181**, 157–159.
den Dunnen, J. T., Moormann, R. J. M., Lubsen, N. H., and Schoenmakers, J. G. G. (1986). *J. Mol. Biol.* **189**, 37–46.
Hunt, C., and Morimoto, R. I. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6455–6459.
Ingolia, T. D., and Craig, E. A. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2360–2364.
Laemmli, U. L. (1970). *Nature (Lond.)* **227**, 680–685.
Marchalonis, J. J., and Weltman, J. K. (1971). *Comp. Biochem. Physiol.* **38B**, 609–625.
Packard, A. (1972). *Biol. Rev.* **47**, 241–304.
Russnak, R. M., Jones, D., and Candido, E. P. M. (1983). *Nucleic Acids Res.* **11**, 3187–3205.
Siezen, R. J. (1981). *FEBS Lett.* **133**, 1–8.
Siezen, R. J., and Shaw, D. C. (1982). *Biochim. Biophys. Acta* **704**, 304–320.
Simpson, R. J., Neuberger, M. R., and Liu, T.-Y. (1976). *J. Biol. Chem.* **251**, 1936–1940.
Wistow, G., Slingsby, C., Blundell, T., Driessen, H., de Jong, W. W., and Bloemendal, H. (1981). *FEBS Lett.* **133**, 9–16.
Wistow, G., Summers, L., and Blundell, T. (1985). *Nature (Lond.)* **315**, 771–773.