

## Sequence Analysis of Four Acidic $\beta$ -Crystallin Subunits of Amphibian Lenses: Phylogenetic Comparison between $\beta$ - and $\gamma$ -Crystallins<sup>1</sup>

Shao-Fan Lu,\* Fu-Ming Pan,† and Shyh-Hong Chiou\*<sup>†,2</sup>

\*Laboratory of Crystallin Research, Institute of Biochemical Sciences, National Taiwan University and †Institute of Biological Chemistry, P.O. Box 23-106, Academia, Taipei, 10098 Taiwan

Received February 26, 1996

$\beta$ -Crystallins composed of the most heterogeneous group of subunit chains among the three major crystallin families of vertebrates, i.e.  $\alpha$ -,  $\beta$ - and  $\gamma$ -Crystallins, are less well understood at the structural and functional levels than the other two. They comprise a multigene family with at least three basic ( $\beta B_{1-3}$ ) and four acidic ( $\beta A_{1-4}$ ) subunit polypeptides. In order to facilitate the determination of the primary sequences of all these ubiquitous crystallin subunits present in all vertebrate species, cDNA mixture was synthesized from the poly(A)<sup>+</sup> mRNA isolated from bullfrog eye lenses. We report here a protocol of Rapid Amplification of cDNA Ends (RACE) was used to amplify cDNAs encoding  $\beta$ -crystallin acidic subunit polypeptides by polymerase chain reaction (PCR). Four complete full-length reading frames with two each of 597 and 648 base pairs, which cover four deduced protein sequences of 198 ( $\beta A1-1$  and  $\beta A1-2$ ) and 215 ( $\beta A3-1$  and  $\beta A3-2$ ) amino acids including the universal initiating methionine, were revealed by nucleotide sequencing. They show about 96–98% sequence similarity among themselves and 76–80%, 80–83% to the homologous  $\beta A1/A3$  crystallins of bovine and human species respectively, revealing the close structural relationship among acidic subunits of all  $\beta$ -crystallins even from remotely related species. In this study a phylogenetic comparison based on amino-acid sequences of various  $\beta A1/A3$  crystallins plus the major basic  $\beta$ -crystallin ( $\beta Bp$ ) and  $\gamma$ -crystallin from different vertebrate species is made using a combination of distance matrix and approximate parsimony methods, which correctly groups these  $\beta A$  crystallin chains together as one family distinct from basic  $\beta$ -crystallins and  $\gamma$ -crystallin and further corroborates the supposition that  $\beta$ - and  $\gamma$ -crystallins form a superfamily with a common ancestry. © 1996 Academic Press, Inc.

The lens crystallins of vertebrates form several complex classes of relatively conserved structural proteins with distant evolutionary relationships [1–3]. We have characterized lens crystallins from various species encompassing one class of invertebrates [4,5] and all five major classes of vertebrates [3,6–9] with the aim of searching for their evolutionary relatedness and phylogenetic relationships.

Three major families of vertebrate crystallins, i.e. those classified as  $\alpha$ -,  $\beta$ -, and  $\gamma$ -crystallins, are found almost in all vertebrates [10].  $\beta$ -Crystallins composed of the most heterogeneous group of subunit chains among the three families are less well understood regarding their structural or functional aspects. They comprise a multigene family of at least seven basic ( $\beta B_{1-3}$ ) and acidic ( $\beta A_{1-4}$ ) polypeptides, with 45–60% sequence identity shared among one another and about 30% with monomeric  $\gamma$ -crystallins [11,12]. Among these polypeptides  $\beta B_p$  (or designated as  $\beta B2$ ) is the predominant subunit chain present in native  $\beta$ -crystallin oligomers [13,14]. We have previously characterized the cDNA sequence of the major  $\beta Bp$  crystallin from the bullfrog lens and constructed a phylogenetic tree based on available homologous  $\beta$ -crystallin sequences [15]. In this report we have further amplified cDNAs constructed from the lenses of the bullfrog employing newer RACE protocols of PCR methodology to aid in the structural analysis of several acidic

<sup>1</sup> The cDNA sequence data of 4 frog acidic  $\beta$ -crystallin subunits,  $\beta A1-1$ ,  $\beta A1-2$ ,  $\beta A3-1$  and  $\beta A3-2$  have been deposited in the EMBL, GenBank and DDBJ nucleotide sequence databases under the accession numbers X87759, X87760, X87761 and X87762, respectively.

<sup>2</sup> Corresponding address: S.-H. Chiou, Institute of Biological Chemistry, Academia, P.O. Box 23-106, Taipei, Taiwan. Fax: (886)-2-3635038; E-mail: SHCHIOUS@bc.ntu.edu.TW.

$\beta$ -crystallin 0polypeptides, circumventing the need for time-consuming conventional protein sequencing due to the presence of a blocked N-terminal group at all  $\beta$ -crystallin subunits [16]. It is deemed essential to determine  $\beta$ -crystallin sequences in order to provide some insights into the structural and genetic basis underlying the complexity of  $\beta$ -crystallin family.

## MATERIALS AND METHODS

*Isolation and protein characterization of frog  $\beta$ -crystallins.* The bullfrogs (*Rana catesbeiana*) used were obtained from a local commercial frog farm. The isolation and purification of frog  $\beta$ -crystallin were essentially as described before [8].  $\beta$ -Crystallin fraction obtained from Fractogel TSK HW-55 gel permeation column was further resolved into its component subunit chains on a reversed-phase  $C_{18}$  column. Amino-acid analysis and N-terminal sequence analysis by automated Edman degradation for each subunit were also carried out as described before [8,17].

*Preparation of total lens crystallin mRNA from frog lenses.* Frog lenses were removed and stored in liquid nitrogen immediately before processing for mRNA isolation. Two deep-frozen lenses from one frog were homogenized and RNA was extracted according to standard procedures [18]. To obtain a full-length crystallin cDNA, poly(A)<sup>+</sup>RNA fraction from total RNA was purified using QuickPrep mRNA preparation kit (Pharmacia, Uppsala, Sweden).

*5'-RACE protocol of PCR amplification.* A RACE protocol [19,20] for PCR amplification was carried out with a 5'-AmpliFINDER RACE Kit of Clontech Laboratories, Inc. (Palo Alto, CA, U.S.A.). cDNA mixture was synthesized from poly(A)<sup>+</sup>mRNA by AMV reverse transcriptase. The synthesized double-stranded RNA-DNA hybrid was then reacted with 6 M NaOH to hydrolyze RNA. The remaining single-stranded DNA was ligated with the "Anchor" fragment. (The Anchor sequence: 5'-CACGAATTC~~ACTATCGATTCTGGAACCTTCAGAGG~~-3') overnight by T<sub>4</sub> RNA ligase through the 3'-end of the cDNA. Following Anchor ligation, a portion of the ligated cDNA was then used as a template for PCR amplification using a primer with the following sequence as the forward primer: (5'-CTGGTTCGGCCACCTCTGAAGGTTCCAGAATCGATAG-3') complementary to the Anchor fragment (underlined) and a  $\beta$ A3/A1 crystallin-specific sequence as the reverse primer: (5'-CTACTGCTGGATGCGCC-3'), covering the 3'-terminal 17 nucleotide segment based on frog and human  $\beta$ A1/A3 crystallins [21,22]. The conditions for PCR reactions were essentially identical to those described previously [23] for frog  $\rho$ -crystallin cDNA amplification. The amplified product was purified using the Wizard PCR Preps DNA Purification System (Promega, Madison, WI) and ligated into pGEM-T vector (1:3 molar ratio of vector to insert). The ligation mixture was incubated at 15°C for 3 h and then transformed into *E. coli* strain JM 109. Transformed cells were mixed with 20  $\mu$ l of 5-bromo-4-chloro-3-indoyl- $\beta$ -D-galactoside (X-gal, 50 mg/mL) and 20  $\mu$ l of isopropylthiogalactoside (IPTG, 100 mM), plated on LB plate (containing 50  $\mu$ g/ml ampicillin) and incubated overnight at 37 °C.

Plasmids purified from positive-clones were then prepared for nucleotide sequencing by the dideoxynucleotide chain-termination method [24]. The DNA sequences were double-checked by an automatic fluorescence-based sequencing of templates amplified by PCR using a model 373A DNA Sequencing System with a *Taq* DyeDeoxy terminator cycle sequencing kit (Applied Biosystems).

*Sequence comparison of  $\beta$ A1/A3 crystallins and various  $\beta$ - and  $\gamma$ -crystallins.* In the comparison and analysis of the deduced amino-acid sequence from the cDNA sequence coding for acidic  $\beta$ A1/A3 crystallin subunits, a multiple-sequence alignment software program (DNASTAR Inc., Madison, WI, U.S.A.) was used for the estimation of sequence homology based on percent sequence similarity.

*Construction of a phylogenetic tree for  $\beta$ - and  $\gamma$ -crystallins.* A software package of LaserGene (for the Apple Macintosh computer) from DNASTAR was used for the estimation of sequence homology based on percent similarity and divergence among different cDNA or protein sequences. Percent divergence is calculated by comparing sequence pairs in relation to the phylogenetic tree. On the other hand the percent similarity is estimated by comparing sequences directly without taking into account of phylogenetic relationships. Phylogenetic trees were then constructed using the algorithm of Hein [25] in the MegAlign programs of the package. It 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 program constructs phylogenetic trees by imposing restrictions based on evolutionary relatedness of the aligned sequences.

## RESULTS AND DISCUSSION

The  $\beta$ - and  $\gamma$ -crystallins are both members of a superfamily of crystallin genes [26], which in contrast to  $\alpha$ -crystallin with chaperone activity appear to lack any detectable functions except existing as abundant structural proteins in most animal eye lenses. We have been interested in the characterization of lens crystallins from the lower submammalian species. It was previously found that during the process of morphological transition from an aquatic tadpole to adult amphibian frog  $\beta$ -crystallin polypeptides appeared to change both in their compositions and quantities [27]. A detailed characterization of these crystallin polypeptides from amphibian lenses should lay a firm

basis to explore the feasibility of using frog lens as a useful model system in the study of structure-function aspects of  $\beta$ -crystallins. This has prompted us to pursue the structural analysis of acidic  $\beta$ -crystallin polypeptides after the recent successful characterization of the major basic  $\beta$ Bp crystallin by PCR cloning and sequencing [15].

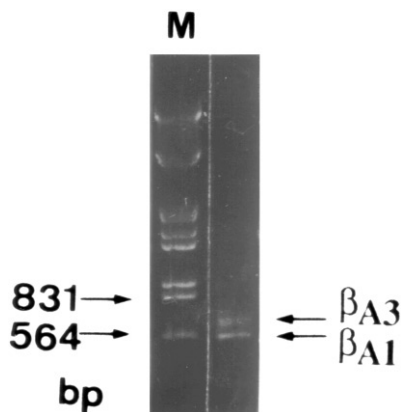
#### *Characterization of $\beta$ -Crystallin Polypeptides from Frog Lenses*

$\beta$ -Crystallin eluted as the second peak from a TSK gel-filtration column with a molecular mass of about 205 kDa. It can be further purified by subjecting the native  $\beta$ -crystallin to ion-exchange chromatography [17] or reversed-phase HPLC [27] to resolve into at least 5 subunit polypeptides. Amino-acid analysis showed that they were extremely similar in composition, and repeated Edman N-terminal analyses by sensitive microsequencing sequencer revealed no amino-terminal residues, indicative of blocked N-termini present in all  $\beta$ -crystallin subunit chains. In this study we have therefore circumvented conventional Edman sequencing by facile PCR cloning and sequencing to solve the structures of these  $\beta$ -crystallin polypeptides.

#### *Sequence Determination of cDNA by 5'-RACE PCR Amplification*

PCR amplification of total lens cDNA mixture with the designed primers corresponding to the N- and C-terminal amino-acid segments of the deduced protein sequences from frog  $\beta$ A1/A3 crystallin nucleotide sequences [21,22] was first employed to generate PCR fragments of  $\beta$ A1/A3 crystallin subunits. In order to validate the determined sequences by PCR based on the N- and C-terminal primers, we have adopted the 5'-end RACE protocol for the PCR amplification of frog  $\beta$ A crystallin cDNAs with a single C-terminal primer to avoid the possible error introduced into the N-terminal sequence due to the use of a prefixed N-terminal primer. By use of 5'-RACE protocol of PCR, the ambiguity of N-terminal segment is therefore cleared.

**Fig. 1** shows the size determination of PCR-amplified cDNA bands coding for  $\beta$ A1/A3 crystallin subunits by 5'-RACE. The amplified PCR products were estimated to be about 650 and 590 bp, encoding two polypeptides of about 195–215 amino-acid residues for  $\beta$ A1 and  $\beta$ A3 respectively. There were about more than 20 positive clones being identified and all of them were shown to consist of four similar yet distinct cDNA sequences by automatic fluorescence-based nucleotide sequencing. It is noteworthy that the amplified cDNA fragments encoding these  $\beta$ A1/A3 crystallin subunits by the conventional method employing N- and C- primers and the newer 5'-RACE using single C-terminal primer possess identical N-terminal sequences. Therefore the fidelity of PCR



**FIG. 1.** Identification and size determination of the PCR-amplified  $\beta$ A1/A3-crystallin cDNAs. Electrophoresis was carried out in 1.2% agarose gel. Lane M, DNA size markers of the *Eco*RI and *Hind*III digestion products of  $\lambda$ DNA, ranging from 564 to 21227 bp; Right Lane indicates the amplified PCR products of about 650 and 590 bp (arrows) encoding frog  $\beta$ A3 and  $\beta$ A1 crystallins respectively.

(A)

```

      10      20      30      40      50      60      70      80      90
ATGCCCCAGATCAACCCCACTTCCTGTTCCT CTGGGACCCCTGGAAGATTACAGTATATGAC CAAGAGAAGCTTTCAGGAAAGAGAATGGAA
M A Q I N P L P V P L G P W K I T V Y D Q E N F Q G K R M E >
      100     110     120     130     140     150     160     170     180
TTCACCTCCTCCTGTGCAAAACATCATGGAG TCGGCGCTTTGATAACATTAGGCTCTGAAA GTGGAATGTGGAGCATGGATTGGATATGAA
F T S S C T N I M E C G F D N I R S L K V E C G A W I G Y E >
      190     200     210     220     230     240     250     260     270
CACACAAGCTTCTCGGTC AACAGTMTTGTCTCTGCTAACCATAAAGAAATCCAAGCTGGTT ATCTTTGAAAAGGAGAATTTTCATTGGGGCC
H T S F C G Q Q F V L E R G E Y P P W D A W S A T N A Y H I >
      280     290     300     310     320     330     340     350     360
GAGCGCTGATGTCTTTCCGCCCAATCTGT TCTGCTAACCATAAAGAAATCCAAGCTGGTT ATCTTTGAAAAGGAGAATTTTCATTGGGGCC
E R L M S F R P I C S A N H K E S K L V I F E K E N F I G R >
      370     380     390     400     410     420     430     440     450
CAGTGGGAATTCGGTGATGACTACCCCTTCT CTACAGGCTATGGGCTGGGAAACAATGAA GTTGGATCCATGAAAGTTCAGTGTGGAGCC
Q W E L R D D Y P S L Q A M G W G N N E V G S M K V Q C G A >
      460     470     480     490     500     510     520     530     540
TGGGTGTGTACCAGTATCCCGGTTATCGT GGCTACCAGTACATTCCTGGAATCCGACCAC CATGAGGTGAATACAAGCACTGGAGAGAG
W V C Y Q Y P G Y R G Y Q Y I L E S D H H G G E Y K H W R E >
      550     560     570     580     590
TGGGGATCCACGCCAGACTTTCCAAATC CAATCTATCCGGCGCATCCAGCAGTAG
W G S H A Q T F Q I Q S I R R I Q Q * >
    
```

(B)

```

      10      20      30      40      50      60      70      80      90
ATGCCCCAGATCAACCCCACTTCCTGTTCCT CTGGGACCCCTGGAAGATTACAGTATATGAC CAAGAGAAGCTTTCAGGAAAGAGAATGGAA
M A Q I N P L P V P L G P W K I T V Y D Q E N F Q G K R M E >
      100     110     120     130     140     150     160     170     180
TTCACCTCCTCCTGTGCAAAACATCATGGAG TCGGCGCTTTGATAACATTAGGCTCTGAAA GTGGAATGTGGAGCATGGATTGGATATGAA
F T S S C A N I M E C G F D N I R S L K V E C G A W I G Y E >
      190     200     210     220     230     240     250     260     270
CACACAAGCTTCTCGGTC AACAGTMTTGTCTCTGCTAACCATAAAGAAATCCAAGCTGGTT ATCTTTGAAAAGGAGAATTTTCATTGGGGCC
H T S F C G Q Q F V L E R G E Y P R W D A W S G S N A Y H I >
      280     290     300     310     320     330     340     350     360
GAGCGCTGATGTCTTTCCGCCCAATCTGT TCTGCTAACCATAAAGAAATCCAAGCTGGTT ATCTTTGAAAAGGAGAATTTTCATTGGGGCC
E R L M S F R P I C S A N H I E S K L V I F E K E N F V G P >
      370     380     390     400     410     420     430     440     450
CAGTGGGAATTCGGTGATGACTACCCCTTCT CTACAGGCTATGGGCTGGGAAACAATGAA GTTGGATCCATGAAAGTTCAGTGTGGAGCC
Q W E L C D D Y P S L Q A M G W G N N E V G S M K V Q C G A >
      460     470     480     490     500     510     520     530     540
TGGGTGTGTACCAGTATCCCGGTTATCGT GGCTACCAGTACATTCCTGGAATCCGACCAC CATGAGGTGAATACAAGCACTGGAGAGAG
W V C Y Q Y P G Y R G Y Q Y I L E S D H H G G E Y K H W R E >
      550     560     570     580     590
TGGGGATCCACGCCAGACTTTCCAAATC CAATCTATCCGGCGCATCCAGCAGTAG
W G S H A Q T F Q I Q S I R R I Q Q * >
    
```

**FIG. 2.** Nucleotide and deduced protein sequences of four frog  $\beta$ A1/A3-crystallins. The nucleotide sequences of 597 (A and B) and 648 base pairs (C and D) are shown above the amino-acid sequences of 198 (A and B) and 215 residues (C and D), which include the translation-initiating methionine as the first amino acid. Asterisks (\*>) indicate the stop codons TAG. cDNA sequences are marked in every 10-nucleotide segment for easy tracing of sequence contents. Amino acids are denoted by one-letter symbols. The 3'-end oligonucleotide segment (17-mer) is the single primer used in the 5'-RACE protocol of PCR amplification to derive the complete cDNA sequences of  $\beta$ A1-1 (A),  $\beta$ A1-2 (B),  $\beta$ A3-1 (C) and  $\beta$ A3-2 (D) crystallins. Note that there is an N-terminal extension of 17 amino acids in  $\beta$ A3-1 and  $\beta$ A3-2 crystallins.

(C)

10	20	30	40	50	60	70	80	90
ATGGAGATCCCACTAGATCAGACTGAGCGA	GAAGACATCACCTCTGAAAAAATGGCCAG	ATCAACCCACTTCCTGTTTCATCTGGGACCC						
M E I P V D Q T E R	E D I T S E K M A Q	I N P L P V H L G P>						
100	110	120	130	140	150	160	170	180
TGGAAGATTACAGTATATGACCAAGAGAAC	TTTCAAGGAAAGAGAATGGAATTCACCTCC	TCCTGTGCAAAACATCATGGAGTGC GGCTTT						
W K I T V Y D Q E N	F Q G K R M E F T S	S C A N I M E C G F>						
190	200	210	220	230	240	250	260	270
GATAACATTAGGTCCTCTGAAAGTGGAAATGT	GGAGCATGGATTGGATATGAACACACAAGC	TTCTGCGGTCAACAGTTTGTCTCTGGAAAGA						
D N I R S L K V E C	G A W I G Y E H T S	F C G Q Q F V L E R>						
280	290	300	310	320	330	340	350	360
GGAGAGTACCCCGCTGGATGCTTGGAGT	GGCAGCAATGCCTATCACATTTGAGCGCCTG	ATGTCCTTCGCCCCAATCTGTTCTGCTAAC						
G E Y P R W D A W S	G S N A Y H I E R L	M S F R P I C S A N>						
370	380	390	400	410	420	430	440	450
CATATAGAATCCAAGCTGGTTATCTTTGAA	AAGGAGAATTTCAATTTGGGCGCCAGCGGGAA	TTGTGTGATGACTACCCCTTCTACAGSCT						
H I E S K L V I F E	K E N F I G R Q R E	L C D D Y P S L Q A>						
460	470	480	490	500	510	520	530	540
ATGGGCTGGGAAAACAATGAAGTTGGATCC	ATGAAAGTTCAGTGTGGAGCCTGGGTGTGT	TACCAGTATCCCGGTTATCGTGGCTACCAG						
M G W G N N E V G S	M K V Q C G A W V C	Y Q Y P G Y R G Y Q>						
550	560	570	580	590	600	610	620	630
TACATTTCTGGAATCCGACCACCATGGAGGT	GAATACAAGCACTGGAGAGAGTGGGGATCC	CACGCCAGACTTTCCAATCCAATCTATC						
Y I L E S D H H G G	E Y K H W R E W G S	H A Q T F Q I Q S I>						
640								
CGGCGCATCCAGCAGTAG								
R R I Q Q *>								

(D)

10	20	30	40	50	60	70	80	90
ATGGAGATCCCACTAGATCAGACTGAGCGA	GAAGACATCACCTCTGAAAAAATGGCCAG	ATCAACCCACTTCCTGTTTCATCTGGGACCC						
M E I P A I Q T E R	E D I T S E K M A Q	I N P L P V P L G P>						
100	110	120	130	140	150	160	170	180
TGGAAGATTACAGTATATGACCAAGAGAAC	TTTCAAGGAAAGAGAATGGAATTCACCTCC	TCCTGTGCAAAACATCATGGAGTGC GGCTTT						
W K I T V Y D Q E N	F Q G K R M E F T S	S C A N I M E C G F>						
190	200	210	220	230	240	250	260	270
GATAACATTAGGTCCTCTGAAAGTGGAAATGT	GGAGCATGGATTGGATATGAACACACAAGC	TTCTGCGGTCAACAGTTTGTCTCTGGAAAGA						
D N I R S L K V E C	G A W I G Y E H T S	F C G Q Q F V L E R>						
280	290	300	310	320	330	340	350	360
GGAGAGTACCCCGCTGGATGCTTGGAGT	GGCAGCAATGCCTATCACATTTGAGCGCCTG	ATGTCCTTCGCCCCAATCTGTTCTGCTAAC						
G E Y P R W D A W S	G S N A Y H I E R L	M S F R P I C S A N>						
370	380	390	400	410	420	430	440	450
CATGAAGAATCCAAGCTGGTTATCTTTGAA	AAGGAGAATTTCAATTTGGGCGCCAGCGGGAA	TTGTGTGATGACTACCCCTTCTACAGSCT						
H E E S K L V I F E	K E N F I G R Q R E	L C D D Y P S L Q A>						
460	470	480	490	500	510	520	530	540
ATGGGCTGGGAAAACAATGAAGTTGGATCC	ATGAAAGTTCAGTGTGGAGCCTGGGTGTGT	TACCAGTATCCCGGTTATCGTGGCTACCAG						
M G W G N N E V G S	M K V Q C G A W V C	Y Q Y P G Y R G Y Q>						
550	560	570	580	590	600	610	620	630
TACATTTCTGGAATCCGACCACCATGGAGGT	GAATACAAGCACTGGAGAGAGTGGGGATCC	CACGCCAGACTTTCCAATCCAATCTATC						
Y I L E S D H H G G	E Y K H W R E W G S	H A Q T F Q I Q S I>						
640								
CGGCGCATCCAGCAGTAG								
R R I Q Q *>								

FIG. 2—Continued

reactions in the amplification of  $\beta$ -crystallin cDNAs is reproducible and reliable in supporting some of our previous studies on frog  $\alpha$ - and  $\gamma$ -crystallin cDNAs employing the PCR protocol of two-primer system [28,29].

The deduced protein sequences together with the nucleotide sequences of frog  $\beta$ A1/A3 crystallin subunits are shown in **Fig. 2 (A)–(D)**. They consist of 597 and 648 nucleotides; which cover four full-length polypeptides of 198 ( $\beta$ A1-1 and  $\beta$ A1-2) and 215 ( $\beta$ A3-1 and  $\beta$ A3-2) amino acids including the initiating methionine. The isoelectric points (pI) for these four frog  $\beta$ A1/A3 crystallin subunits calculated from the deduced protein sequences are 6.17, 5.77, 5.57 and 5.45 respectively, in agreement with the common nomenclature in the identification of these  $\beta$ -crystallin as acidic  $\beta$ -crystallin subunits.

### Structural Comparison of Four Frog $\beta$ A1/A3 and the Homologous Crystallin Subunits from Other Vertebrate Species by Multiple-Sequence Alignment

**Fig. 3(A)** shows the optimal alignment of the four deduced protein sequences from cDNAs for  $\beta$ A1/A3 subunits. It is of interest to find that they are highly homologous to each other as judged

(A)



**FIG. 3.** Multiple sequence alignment and sequence comparison among bullfrog  $\beta$ A1/A3-crystallins (A) and homologous  $\beta$ A3/A1-crystallins and bullfrog  $\gamma$ M1-1 crystallin [29] (B). Sequence alignment was carried out in the DNASTAR program using the published sequences of  $\beta$ A3/A1 crystallins from *R. temporaria* [22], chicken [32], bovine [12] and human [21]. The gaps were introduced for optimal alignment and maximum homology for the sequences. The amino-acid residues located in the highlighted regions are conserved based on comparison of the sequences. Amino acid residues are denoted by one-letter symbols. Note that in bovine  $\beta$ A3/A1-crystallin there is a missing segment in the middle region due to incomplete protein sequencing.

(B)

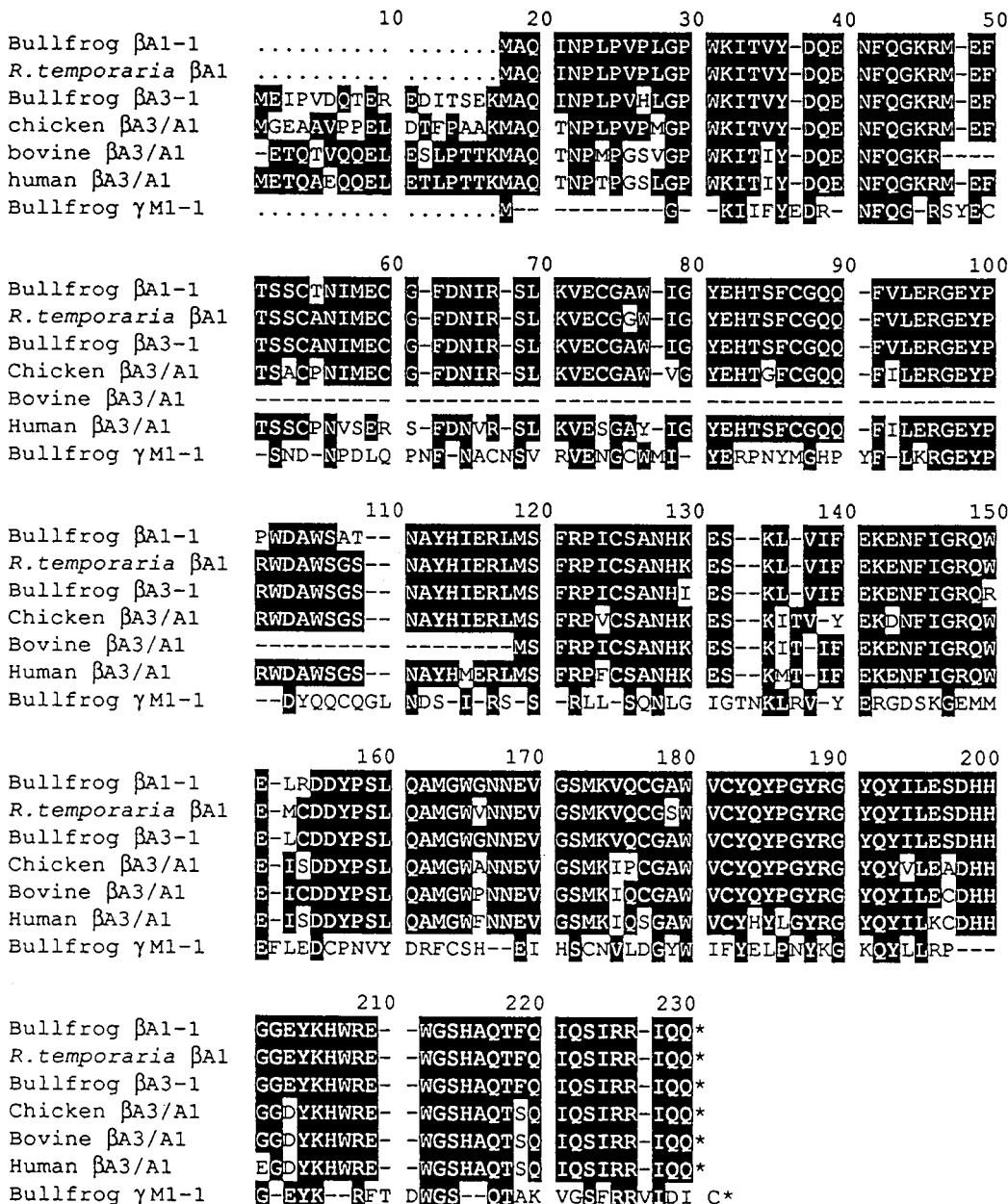


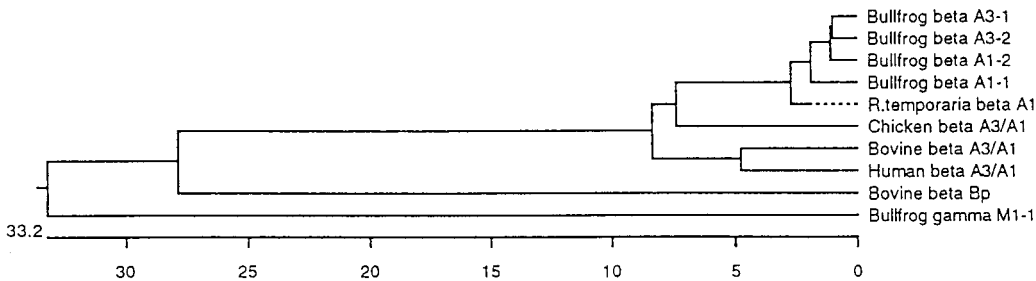
FIG. 3—Continued

by percents sequence homology of 96% and 98% shared in the pairs of  $\beta$ A1-1/ $\beta$ A1-2 and  $\beta$ A3-1/ $\beta$ A3-2 respectively.  $\beta$ A1-1/ $\beta$ A1-2 and  $\beta$ A3-1/ $\beta$ A3-2 are also very similar regarding their sequences with the exception that there is an N-terminal extension of 17 amino acid residues in the  $\beta$ A3-1 and  $\beta$ A3-2. It has been reported that  $\beta$ A1- and  $\beta$ A3-crystallins are synthesized by alternate translation initiation from a common mRNA of the mammalian  $\beta$ A1/A3 crystallin gene for human [21], mouse [30] and bovine [31] species. In this report we further strengthened and extended the

(A)

		Percent Similarity											
		1	2	3	4	5	6	7	8	9	10		
Percent Divergence	1	■	88.6	81.4	80.0	80.5	83.3	83.3	83.8	29.4	21.5	1	Human beta A3/A1
	2	9.4	■	81.2	79.9	79.9	77.2	75.8	77.2	26.8	17.4	2	Bovine beta A3/A1
	3	18.6	16.8	■	81.9	82.8	87.4	87.4	87.9	32.8	23.7	3	Chicken beta A3/A1
	4	20.0	18.1	18.1	■	98.1	96.0	98.0	96.5	30.9	23.2	4	Bullfrog beta A3-1
	5	19.5	18.1	17.2	1.9	■	96.5	98.0	97.0	30.9	21.5	5	Bullfrog beta A3-2
	6	16.7	11.4	12.6	4.0	3.5	■	96.0	95.5	31.8	23.7	6	Bullfrog beta A1-1
	7	16.7	12.9	12.6	2.0	2.0	4.0	■	96.5	31.8	23.2	7	Bullfrog beta A1-2
	8	16.2	11.4	12.1	3.5	3.0	4.5	3.5	■	32.3	24.3	8	R.temporaria beta A1
	9	58.0	53.2	55.9	56.4	55.9	55.3	54.2	54.2	■	28.2	9	Bovine beta Bp
	10	68.4	67.0	65.5	66.1	66.1	66.5	65.9	65.9	66.1	■	10	Bullfrog gamma M1-1
		1	2	3	4	5	6	7	8	9	10		

(B)



**FIG. 4. (A)** Pair-wise comparison of the percentage similarity and divergence for amino-acid sequences using the multiple-sequence alignment program *LaserGene* by DNASTAR Inc. (Madison, WI, USA). The percentage sequence 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. The sequences of  $\beta Bp$  and  $\gamma M1-1$  were taken from bovine [15], bullfrog [29] and  $\beta A3/A1$  crystallins from *R. temporaria* [22], chicken [32], bovine [12] and human [21]. **(B)** Phylogenetic tree of eight  $\beta A3/A1$  and bovine  $\beta Bp$  plus bullfrog  $\gamma M1-1$  crystallin sequences. The trees were constructed based on the percentage divergence between protein sequences using a combination of distance matrix and approximate parsimony methods according to the phylogeny generation program of Hein (1990) [25]. This program carries out multiple alignment by imposing restrictions based on the evolutionary relatedness of the aligned sequences. The length of each pair of branches represents the estimated sequence distance between aligned pairs. The scale beneath the tree measures the evolutionary distance between sequences (in millions of years). The dotted line shown for *R. temporaria* points to the fact that the sequence distance is not proportional to the scale.

supposition that translation of  $\beta A1/A3$  crystallin subunits commencing at two different initiation codons operates also at the lower vertebrate species.

In **Fig. 3(B)** we have chosen two representative bullfrog  $\beta A1-1$  and  $\beta A3-1$  sequences for a sequence alignment and homology comparison with those sequences reported previously for the homologous  $\beta A3/A1$  crystallins from another common frog [22], chicken [32], bovine [12] and human [21] plus one bullfrog  $\gamma M1-1$ . It is clearly seen that with the exception of bullfrog  $\gamma M1-1$  crystallin which shows only 21–24% sequence identity with bullfrog  $\beta A3/A1$  crystallins, the sequence similarity among various  $\beta B A1/A3$  crystallins from amphibian to mammalian species is very high (76–80% homology between bullfrog and bovine and 82–88% between bullfrog and chicken) even from different classes of vertebrates, again underlining the close relatedness of this



conserved  $\beta$ A1/A3 crystallin family. On the other hand, bullfrog  $\gamma$  crystallins [29] show only medial sequence similarity (<25%) to frog  $\beta$ A3/A1 crystallins despite the fact that  $\beta$ - and  $\gamma$ -crystallins have been suggested to form a  $\beta/\gamma$  superfamily based on their partial sequence homology [11,12]. It is to be noted that the N-terminal extension which is present in  $\beta$ A3-1/ $\beta$ A3-2 and not in  $\beta$ A1-1/ $\beta$ A1-2 or  $\gamma$ -crystallin shows much greater variation among different species. Both  $\beta$ - and  $\gamma$ -crystallins are proposed to possess a conserved two-domain structure [13,14] and N-terminal extension is believed to be involved in the interactions among different  $\beta$ -crystallin subunits to form higher  $\beta$ -crystallin oligomers.

### *Construction of a Phylogenetic Tree Based on Protein Sequences*

In the comparative analysis of sequence similarity based on deduced amino-acid sequences using multiple-sequence alignment program (DNASTAR Inc.), it was found that  $\beta$ A1/A3 crystallins of the bullfrog show 76–87% sequence similarity (**Fig. 4A**) to the homologous  $\beta$ A1/A3 crystallins of human, bovine and chicken lenses, which is higher than DNA sequence homology detected among these species. We have therefore constructed a phylogenetic tree (**Fig. 4B**) based on the percentage divergence calculated from the pair-wise comparison of protein sequences of these related crystallins using the algorithm of Hein [25], 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. In our previous studies [15,28,29] we have found this algorithm is very useful to align highly evolved gene families and their corresponding protein sequences that exhibit clear evolutionary relationships as in the cases of  $\alpha$ -,  $\beta$ - and  $\gamma$ -crystallin families. The phylogenetic tree constructed from the comparison of these sequences exemplifies the close relatedness for eight homologous acidic  $\beta$ A1/A3 crystallins which shared a common ancestral origin with the basic  $\beta$ Bp crystallin. There is a more remote relationship between  $\beta$ A1/A3,  $\beta$ Bp and  $\gamma$ -crystallin even though they are presumably derived from one common ancestral protein with a structure similar to the  $\beta/\gamma$  superfamily. The identification of a separate group of  $\beta$ A1/A3 crystallins in the amphibian class distinct from  $\beta$ B crystallins attests to the complexity in the evolution of  $\beta$ -crystallin family.

### ACKNOWLEDGMENTS

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

### REFERENCES

1. de Jong, W. W. (1981) in *Molecular and Cellular Biology of the Eye Lens* (Bloemendal, H., Ed.), pp. 221–278, Wiley, New York.
2. Bloemendal, H. (1982) *CRC Crit. Rev. Biochem.* **12**, 1–38.
3. Chiou, S.-H. (1986) *FEBS Lett.* **201**, 69–73.
4. Chiou, S.-H. and Bunn, H. F. (1981) *Invest. Ophthalmol. Vis. Sci. ARVO Suppl.* **20**, p. 138.
5. Chiou, S.-H. (1984) *J. Biochem.* **95**, 75–82.
6. Chiou, S.-H., S.-W. Chen, and T.-B. Lo (1986) *FEBS Lett.* **209**, 107–110.
7. Chiou, S.-H., Chang, T., Chang, W.-C., Kuo, J., and Lo, T.-B. (1986) *Biochim. Biophys. Acta* **871**, 324–328.
8. Chiou, S.-H. (1987) *Int. J. Peptide Protein Res.* **30**, 108–116.
9. Chiou, S.-H., Chang, W.-P., Lo, C.-H., and Chen, S.-W. (1987) *FEBS Lett.* **221**, 134–138.
10. Harding, J. J. and Dille, K. J. (1976) *Exp. Eye Res.* **22**, 1–73.
11. Berbers, G. A. M., Boermann, O. C., Bloemendal, H., de Jong, W. W. (1982) *Eur. J. Biochem.* **128**, 495–502.
12. Berbers, G. A. M., Hoekman, W. A., Bloemendal, H., de Jong, W. W., Kleinschmidt, T., and Braunitzer, G. (1984) *Eur. J. Biochem.* **139**, 467–479.
13. Berbers, G. A. M., Brans, A., Hoekman, W. A., Slingsby, C., Bloemendal, H., and de Jong, W. W. (1983) *Biochim. Biophys. Acta* **748**, 213–219.
14. Bax, B., Lapatto, R., Nalini, V., Driessen, H., Lindley, P. F., Mahadevan, D., Blundell, T. L., and Slingsby, C. (1990) *Nature* **347**, 776–780.
15. Pan, F.-M., Chang, W.-C., Lu, S.-F., Hsu, A.-L., and Chiou, S.-H. (1995) *Biochem. Biophys. Res. Commun.* **217**, 940–949.

16. Groenen, P. J. T. A., Merck, K. B., de Jong, W. W., and Bloemendal, H. (1994) *Eur. J. Biochem.* **225**, 1–19.
17. Chiou, S.-H., Chang, W.-P., Chen, S.-W., and Lo, C.-H. (1988) *Int. J. Peptide Protein Res.* **31**, 335–338.
18. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, NY.
19. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
20. Edwards, J. B. D. M., Delort, J., and Mallet, J. (1991) *Nucleic Acids Res.* **19**, 5227–5232.
21. Hogg, D., Tsui, L.-C., Gorin, M. B., and Breitman, M. L. (1986) *J. Biol. Chem.* **261**, 12420–12427.
22. Luchin, S. V., Zinovieva, R. D., Tomarev, S. I., Dolgilevich, S. M., Gause Jr. G. G., Bax, B., Driessen, H., and Blundell, T. L. (1987) *Biochim. Biophys. Acta* **916**, 163–171.
23. Lu, S.-F., Pan, F.-M., and Chiou, S.-H. (1995) *Biochem. Biophys. Res. Commun.* **214**, 1079–1088.
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
25. Hein, J. (1990) *Methods Enzymol.* **183**, 626–645.
26. Driessen, H. P. C., Herbrink, P., Bloemendal, H., and de Jong, W. W. (1981) *Eur. J. Biochem.* **121**, 83–91.
27. Jiang, Y.-J., Chiou, S.-H., and Chang, W.-C. (1989) *Biochem. Biophys. Res. Commun.* **164**, 1423–1430.
28. Lu, S.-F., Pan, F.-M., and Chiou, S.-H. (1995) *Biochem. Biophys. Res. Commun.* **210**, 974–981.
29. Lu, S.-F., Pan, F.-M., and Chiou, S.-H. (1996) *J. Protein Chem.* **15**, 103–113.
30. Inana, G., Piatigorsky, J., Norman, B., Slingsby, C., and Blundell, T. (1983) *Nature* **302**, 310–315.
31. Gorin, M. B., and Horwitz, J. (1984) *Curr. Eye Res.* **3**, 939–948.
32. McDermott, J. B., Peterson, C. A., and Piatigorsky, J. (1992) *Gene* **117**, 193–200.